Assessment Testing: Not a Measure of Intelligence, But Certainly the Means to Marginalize Political Participation and Economic Opportunity

[I]n contemporary societies the trappings of science are readily used, in good faith, to produce disastrously false results. These results become the stock-in-trade of vested interests. When doubts are uttered, money and prestige are threatened, and indeed all of society is shaken, at least in its easy assumptions.
Banesh Hoffmann

Education Is Not the Learning of Facts, But the Training of the Mind to Think. Attributed to Albert Einstein

The human brain should be used for processing, not storage.
Thomas A. Kelly

Bill Murrin
2017[1]

Applied Education Foundation
Promoting Education in the Useful Arts & Sciences

[1] In this and all my other essays, I will periodically add applicable supplemental information as new information becomes available. Therefore, this published year refers to its first release to the public.

Table of Contents

Introduction

The United States culture of testing is, in large measure, an evil that bars a majority of citizens from full participation in society. The testing methodology, employed by educational bureaucracies, does not

measure intelligence, but, rather, measures memory and recall abilities. People without these talents are politically and economically marginalized since the primary path to credentials is through the highly optimized educational establishment's rules[2] dictated by accreditation organizations, which the testing industry echoes. Therefore, based on the arbitrary accreditation and testing industries controls, the future of every citizen is, more often than not, determined by their arbitrary demands. This is a result of the institutionalization of education; and institutions evolve into an evil with some social utility. This is made evident by Einstein's opinion on knowing facts versus developing the ability to reason.

The Einstein quote – or maybe it's a paraphrase – on the title page sums up the difference between the current state of education versus its earlier structure and intentions. When an individual contacted *Quote Investigator* and stated, "A learner may accumulate a large number of miscellaneous pieces of information without achieving an integrated understanding and without acquiring an ability to use the material intelligently." Then he asked if the statement "Education is not the learning of facts, but the training of minds to think" is attributable to Albert Einstein. They researched the source of this statement and while they didn't find an exact quote to this effect, they found quotes that were close and which offer a bit more meat to chew on. They found:

> In 1921 Albert Einstein visited Boston, Massachusetts. At that time, a questionnaire constructed by the inventor and research laboratory pioneer Thomas A. Edison was circulating. Edison used his controversial questionnaire to screen job applicants, but Einstein was unimpressed by some of the queries. For example, "The New York Times" reported on Einstein's reaction to one question …: He was asked, "What is the speed of sound?" He could not say off-hand, he replied. "I don't know. I don't burden my memory with such facts that I can easily find in any textbook."
>
> Einstein's response … fit the theme of the quotation because he deemphasized the value of simply memorizing facts. A longer description of this episode was presented in the biography "Einstein: His Life and Times" by Philipp Frank, 1947. A strong match for the quotation was included in the following passage:
>
> "Nor did he agree with Edison's opinion on the uselessness of college education. He remarked: 'It is not so very important for a person to learn facts. For that, he does not really need a college. He can learn them from books. The value of an education in a liberal arts college is not the learning of many facts but the training of the mind to think, something that cannot be learned from textbooks.'
>
> "'… I have heard one college professor … frequently tell his students that they were not in college to learn facts, but to train their minds to think logically. He was right, but that training should have been started many years before. The man who expects to learn to think after he has reached college is the man who "flunks" or just barely hangs on.'"
>
> Einstein minimized the value of learning motley textbook facts, and highlighted the value of training the mind.
>
> … In conclusion, the quotation under examination was a streamlined/simplified version….

[2] See my essay *How Much Education is Really Necessary*, "The Optimization of College for Certain Abilities," pages 22-27.

Quote Investigator's research provides different angles to get the point across that the accumulation of facts and data, currently passing as education's objective, is largely a waste of time. If students aren't taught how to think, education should be condensed to a few years of study to the level of competency in reading, writing and arithmetic (as was the case before the bureaucratization of education that manifest in the latter half of the 19th century); and then provide them access to the world of knowledge through libraries and the internet so that they or parents may educate themselves in what they need and are interested in.

* * *

Guskey (2015) has invested a great deal of time and effort in understanding the problems with our assessment methodologies. While he provides severe criticisms of them with thoughtful alternatives to consider, he does not get to the root of the problem in looking at what is taught, how it is taught, and why it is taught. However, he provides an excellent analysis that must be explored to begin the process of repairing a failed system. To start, he quotes a 1913 master's thesis written by I. E. Finkelstein, a Cornell student:

> When we consider the practically universal use in all educational institutions of a system of marks, whether numbers or letters, to indicate scholastic attainment of the pupils or students in these institutions, and when we remember how very great stress is laid by teacher and pupils alike upon these marks as real measures or indicators of attainment, we can but be astonished at the blind faith that has been felt in the reliability of the marking system. School administrators have been using with confidence an absolutely uncalibrated instrument. Only within a very few years have serious attempts been made to scrutinize the theory of marking or to test by statistical and experimental procedure the degree of precision that could be expected in its use.
>
> … Few teachers stop to consider what the marking system under which they work really implies; that the variability in the marks given for the same subject and to the same pupils by different instructors is so great as frequently to work real injustice to students.

Guskey then points out that few would argue with Finkelstein's observations as they relate to our current education system. He analyzes the reasons no change has occurred in all this time:

> Education scholars have recognized these problems and noted their seriousness for well over a century. So then why, if we've known about these problems for so long, have we not found a solution? With all that we have learned about education over the past hundred years, why have grading and reporting continued essentially unchanged?
>
> Oddly, part of the explanation may rest in the seriousness of the consequences attached to grades and the fear that changing grading might disrupt the traditions that yield those consequences. For more than a century, grades have remained the primary indicator of how well students performed in school and the basis for making important decisions about students. Grades largely determine whether or not students get promoted from one grade level to the next. They are used for determining honor roll status, membership in honor societies, and enrollment in advanced classes. High grades are required for admission to selective colleges …, and low grades typically

> are the first indicator of potential learning … disabilities that may require special intervention. Because the relationship between grades and these consequences is so powerful and so direct, changing the way we grade could disrupt that relationship and confound crucial decision-making procedures.
>
> In addition, over the years, grades have served well the sorting and selecting purposes of schools. Historically, schools have functioned to identify, and in some cases to accentuate, the differences among students. Those few students with innate talent and ability generally received high grades, continued in school, went on to college, and became leaders in professions, government, business, and industry. The majority of students received more mediocre grades and learned only what was needed to gain meaningful employment and contribute in an industrial society. Schools ranked students on the basis of their grades to make this selection process easier for colleges as well as for potential employers.
>
> Today, however, our society has changed. We no longer are educating students for an industrial society and its associated jobs. Modern technology has made most of those jobs obsolete [*I would say "many" not "most"*]. Instead we must educate students for a continuously evolving information society [*this is only partially true*] that demands flexibility, creativity, and initiative. We cannot afford to have only a few students learn well and reach high levels of competence or proficiency. To remain competitive in our increasingly complex and technologically sophisticated world, we need all students to learn well. Educators must do their best to help every student develop the advanced knowledge and 21st century skills necessary to enter college or be ready to begin a career that requires creativity, collaboration, advanced reasoning, and problem solving. **Instead of being concerned with selecting the talented few, we must be committed to developing the unique talents of all students**. (pp. 3-4) (Emphasis and comments added.)

I feel that Guskey's summary of the situation is outstanding – though he did miss the mark as it relates to how terribly the public system was designed for citizens even for an industrial society – and is a good point of reference to start the dialogue for change as it relates to the means to determine competency in subjects that are relevant to individuals.

One issue that requires further investigation is the trustworthiness of assessment results that rely exclusively on 2-dimensional interaction. That is: What is the value of tests performed on paper or a computer screen versus those performed in 3-dimensional, real-world activities? It's one thing to ask a question, expecting a verbal or written answer, based on complex reasoning problems, it is quite another thing to face it in the real world and perhaps even perform the activity, if practicable to do so. It cannot be doubted that in many cases, those who perform well in the 2-dimensional realm will perform quite differently in the 3-dimensional realm.

It is not the purpose of this paper to identify and choose – i.e. prescribe – the "best method" to assess educational outcomes. Rather, the purpose of this essay is to try to identify some of the problems with current testing regimes and discuss alternative approaches, not only for assessment, but also regarding what is taught and how it is taught. There are three issues then that need to be researched: 1) the elimination of ranking, replacing it with competency based principles, and summarized in individual portfolios; 2) the identification of revised educational subject matter that is coupled with the various talents people possess (think of multiple intelligences) and individual life goals; and 3) establish

assessment methods, extracted from the private sector, for various educational programs that will be offered to a diverse talent pool for diverse career pathways that make up a contemporary society. This essay will not address these issues in an outline form. It will, however, address them at various points throughout.

Subjects need to be taught and assessed in ways that will not discourage learning. Our almost exclusive dependence on the ability to memorize and recall abstract disconnected information for tests is completely unacceptable and reveals the cause of the problem with the entire assessment paradigm. Such abilities do not reveal intelligence but rather demonstrate computer-like abilities.

Other cognitive abilities, in large part, are dismissed as irrelevant, even other abilities that provide the means for a society to function. The few occupations that are highly screened for IQ are service-oriented professions. They are typically not the producers and most are therefore of secondary importance to society, though they are the primary group we invest education resources into. There is something very wrong with this scenario.

I believe Dierking (2017) quoting Terrel Rhodes, vice president for the Association of American Colleges & Universities (AAC&U), sums up the inadequacies of our current assessment regime very well:

> [A] test of multiple-choice questions about a subject can measure how much a student knows about that field. But it does not show how well a student can connect that information to the real world. And, Rhodes says, it fails to show how well a student can think about problems in different ways.
>
> Also, most of the test results are not used to improve student learning. Their main purpose is to inform government officials who provide funding to schools, as well as accrediting agencies.

Spencer (1860) has this to say about what is taught for examinations and its relevance to life:

> Examinations being once passed, books are laid aside; the greater part of what has been acquired, being unorganized, soon drops out of recollection; what remains is mostly inert – the art of applying knowledge not having been cultivated; and there is but little power either of accurate observation or independent thinking. To all which add, that while much of the information gained is of relatively small value, an immense mass of information of transcendent value is entirely passed over. (p. 53)

Spencer also provides:

> [T]he mind, like the body, cannot assimilate beyond a certain rate; and if you ply it with facts faster than it can assimilate them, they are very soon rejected again: they do not become permanently built into the intellectual fabric; but fall out of recollection after the passing of the examination for which they were got up. It is a mistake too, because it tends to make study distasteful … [and] an aversion to books. … It is a mistake, also, inasmuch as it assumes that the acquisition of knowledge is everything; and forgets that a much more important matter is the organization of knowledge…. Just as Humboldt remarks respecting the progress of intelligence

> in general, that "the interpretation of nature is obscured when the description languishes under too great an accumulation of insulated facts;" so it may be remarked, respecting the progress of individual intelligence, that the mind is overburdened and hampered by an excess of ill-digested information. It is not the knowledge stored up as intellectual fat which is of value; but that which is turned into intellectual muscle. (pp. 300-02)

Kallendorf (2002) provides, **"But it is also very true that those who are keener in mind are less strong in memory, and those who grasp things quickly retain fewer of them."** (p. 63) (Emphasis added) It is interesting that no individual possesses all the talents/intelligences humans possess. It is evident that where an individual possesses one talent at a high level, another or other talents are proportionately reduced. Kallendorf's quote reveals this tendency, but it is something academe seems largely unaware of.

What I've found in my own strengths and weaknesses supports Kallendorf's position. I have difficulty in recalling data – such as names, dates, events, and words – quickly, though if given time, I will find them in various places besides my memory. I see such data as superficial and readily accessible in various places, such as my cell phone that provides access to all the knowledge and data that can be found on the Internet. On the other hand, I am very quick at recalling principles; principles are second nature to me. But then, with Internet accessibility, people with my talent are at the advantage. It appears the days when memory and recall abilities dominated our society are at an end – they have become obsolete.

To conclude this introduction, the Massachusetts Business Alliance for Education provides a report that offers insight into assessment testing that is representative of the entire country and stipulates the reason why a paradigm shift in education is in order.

> In a recent survey conducted for the Massachusetts Business Alliance for Education (MBAE), employers representing a wide range of industries across Massachusetts told us that our education system is out of sync with their expectations and needs. Sixty-nine percent of employers report having difficulty hiring people with the right competencies to fill open positions. **Many cited a lack of applied, real world skills as the problem and expressed concerns about current standardized tests that don't measure abilities needed for success outside of school**. (p. 1) (Emphasis added)

Our Current Assessment Dilemma is Over a Century Old

Guskey (2015) offers a brief historical account of grading and reporting:

> Grading and reporting are relatively recent phenomena in education. In fact, prior to 1850, grading and reporting were virtually unknown in schools in the United States. Throughout much of the 19th century, most schools grouped students of all ages and backgrounds together with a single teacher in one-room schoolhouses, and few students went beyond the elementary level. The teacher reported students' learning progress orally to parents, usually during visits to students' homes.
>
> As the number of students increased in the late 1800s, schools began to group students in grade levels according to their age, and new ideas about curriculum and teaching methods were tried.

> One of these new ideas was the use of formal progress evaluations of students' achievement in school. In these evaluations, teachers simply wrote down the skills each student had mastered and those on which additional work was needed. This was done primarily for the students' benefit, since they were not permitted to move on to the next level until they demonstrated their mastery of the current one.[3] It was also the earliest example of a narrative report card.
>
> With the passage of compulsory school attendance laws at the elementary level during the late 19th and early 20th centuries, the number of students entering high schools increased rapidly. … As a result, subject-area instruction in high schools became increasingly specific, and student populations became more diverse. While elementary teachers continued to use written descriptions and narrative reports to document student learning, high school teachers began using percentages and other similar markings to certify students' accomplishments in different subject areas. This was the beginning of the grading and reporting systems that exist today.
>
> The shift to percentage grades was gradual, and few American educators questioned it. The practice seemed a natural result of the increased demands on high school teachers, who now faced classrooms with growing numbers of students. But in 1912, a study by two Wisconsin researchers seriously challenged the reliability of percentage grades as accurate indicators of students' achievement. (pp. 24-25)

Guskey then provides details of the research that demonstrated the incredible and extreme variation in grading teachers give to assignments. It proved how subjective and arbitrary grading tends to be with catastrophic consequences for individuals who receive the short end of the stick due to random forces beyond their control.

Quick (1894) laments the degradation of education in the late 19th century and reflects on Montaigne's concepts of what a good education entails.

> [Montaigne] thought virtue and judgment were the main things to be cared for…. The one thing gained, or supposed to be gained in the public schools was the art of living, and this art, though it does not demand heroic virtue, requires at least prudence and self-control. … So the education *out of school* was in his eyes of more value than the education in school. And this was acknowledged also in our public schools: "It is not the Latin and Greek they learn or don't learn that we consider so important," the masters used to say, "but it is the tone of the school and the discipline of the games." But of late years this virtual agreement with Montaigne has been broken up. School work is no longer mere employment, but it is done under pressure, and with penalties.
>
> … What has produced this great change? It is due mainly to two causes:
>
> The pressure put on the young to attain classical knowledge was relaxed when it was thought that they could get through life very well without this knowledge. But in these days new knowledge

[3] Of course, this is fundamentally what competency-based education is all about and we should certainly return to the principle of requiring students to master competency before moving to the next level for literacy and numeracy foundational requirements. Not all subjects need such demands as literacy and numeracy; therefore all subjects must lend support to their mastery.

> has awakened a new enthusiasm. The knowledge of science promises such great advantages that the latest reformers … seem to make the well-being of the grown person depend mainly on the amount of scientific knowledge he stored up in his youth. This is the first cause of educational pressure.
>
> The second and more urgent cause is the rapid development of our system of examinations. Everybody's educational status is now settled by the examiner, a [monarch] whose influence has brought back in a very malignant form all the evils of which Montaigne complains. Do what we will, the faculty chiefly exercised in preparing for ordinary examinations is the "carrying memory." So the acquisition of knowledge – mere memory or examination knowledge – has again come to be regarded as the one thing needful in education, and there is great danger of everything else being neglected for it. … A great change has come over our public schools. The amount of work required from the boys is far greater than it used to be and masters again measure their success by the amount of knowledge the average boy takes away with him. It seems to me high time that another Montaigne arose to protest that a man's intellectual life does not consist in the number of things he remembers, and that his true life is not his intellectual life only, but embraces his power of will and action and his love of what is noble and right. … In these days of science and examinations does there not seem some danger lest knowledge should prove the sole survivor? May not Knowledge, like another Cain, raise its hand against its brethren…? This is perhaps the great danger of our time, a danger especially felt in education.

Quick's comparison between Knowledge and Cain proved prophetic when we consider how, in the name of the greatest good for the greatest amount of people, science was used to annihilate many millions of people during both world wars, which would have been less likely to have occurred if Montaigne's idea of education, being based on virtue and judgment, had endured. Knowledge, without virtuous principles underpinning it, will inevitably be used to further the advantage of those who possess the superior position. Just consider how abusive educators can be of those who they feel "can't cut the mustard," which is due to a lack of virtue found in academia and which explains academia's disregard for the educational needs of those who they have no respect for.

An example of the complimentary method of teaching subject matter with an underpinning of virtuous principles can be found in the 1945 movie *A Tree Grows in Brooklyn*, depicting a very poor family trying to make ends meet around the turn of the last century – a true tear-jerker. "The story centers on the main character, Francie Nolan, whose father is a silver-tongued charmer who dazzles Francie with his fantastic dreams. Because of his irresponsibility and heavy drinking, it falls to his wife to hold the family together." In the 8th grade Christmas school-room scene, Francie's teacher, not knowing anything about the father's problems, tries to teach Francie how to be a creative writer coupled with a moral perspective. In so doing, the teacher reveals to the audience the problems with Francie's father who, unintentionally and perhaps innocently, hurts himself along with the ones he truly and deeply loves. The life lesson the teacher provides Francie – i.e. wisdom –, which is something the father was never taught, is of greater value than the knowledge Francie was receiving in composition. Montaigne would certainly have agreed with this.

To digress for a moment: This story reflects on the development of virtue and responsibility in individuals rather than addressing "social justice" – an abstract and highly controversial political concept that maliciously pits one sector of society against another – that is currently taught in public

schools and is the antithesis of teaching for virtue. That is, in the name of virtuous ends, the new religion, "social justice," uses destructive means to demand "virtuous" compliance. It appears the "church" is once again in control of the state.

Next let us consider Quick's analysis of John Locke, where Quick states:

> Opposed to [Locke's position on education] we have the schoolmaster's ideal which is governed by examinations. According to this ideal the object of the school course is to give certain "knowledge," linguistic and other, and to fix it in the memory in such a manner that it can be displayed on the day of examination. "Knowledge" of this kind often makes no demand whatever on the reasoning faculty, or indeed on any faculty but that of remembering and reproducing what the learner has been told; in extreme cases the memory of mere sounds or symbols suffices.

Quick then provides a quote further criticizing the prevailing tendency in his day toward a dependence on accumulated data for regurgitation at periodic examination intervals in order to judge and categorize students to further the interests of the educational establishment:

> "Examinations directed, as the paper examinations of the numerous examining boards now flourishing are directed, to finding out what the pupil *knows*, have the effect of concentrating the teacher's effort upon the least important part of his function." Mark Pattison in *N. Quart. M.*, Jan. 1880.

Quick further quotes Lord Armstrong who wrote "The Cry for Useless Knowledge" (*Nineteenth Century Magazine*, Nov. 1888):

> The aphorism that knowledge is power is so constantly used by educational enthusiasts that it may almost be regarded as the motto of the party. But the first essential of a motto is that it be true, and it is certainly not true that knowledge is the same as power, seeing that it is only an aid to power. … Knowledge is not even an aid to power in all cases, seeing that useless knowledge, which is no uncommon article in our popular schools, has no relation to power. The true source of power is the originative action of the mind which we see exhibited in the daily incidents of life, as well as in matters of great importance. … A man's success in life depends incomparably more upon his capacities for useful action than upon his acquirements in knowledge, and the education of the young should therefore be directed to the development of faculties and valuable qualities rather than to the acquisition knowledge. … Men of capacity and possessing qualities for useful action are at a premium all over the world, while men of mere education are a at a deplorable discount.
>
> There is a great tendency in the scholastic world to underrate the value and potency of self-education, which commences on leaving school and endures all through life.

> I deprecate plunging into doubtful and costly schemes of instruction, led on by the *ignis fatuus*[4] that 'knowledge is power.' For where natural capacity is wasted in attaining knowledge, it would be *truer* to say that knowledge is weakness. (pp. 77-79)

Quick may have been somewhat stuck in a bygone day when extended education was the province of the wealthier classes in society, which comprised a relatively small percentage of the population. However, those bygone years Quick lamented the loss of, had a better understanding of examining students' learning. Educators, and the community in general, were not as caught up in the miniscule details of what students memorized, such as we have today and which currently marginalizes a majority of student populations. They were more concerned with the principles that encompass life since this is all that really matters in a general education. Once a career is chosen, and a student pursues education in a chosen field, the miniscule details become important to that field of endeavor.

Given this perspective, testing methods we currently employ, such as multiple choice and true-false structured methods designed by so-called scientists and bureaucratic institutional type people, were not used prior to the latter half of the 19th century. Educators wanted to see students' written work. Teachers had a great deal of autonomy (as currently seen in Finland) and were able to ignore miniscule mistakes in order to focus on the content of students' work.

The Tyranny of Testing

Hoffmann's[5] work, *The Tyranny of Testing* (1964), was a work that, in its day, created a seismic shift in education. However, due to vested interests in the educational establishment – that is, the testing industry in particular – it was eventually marginalized and is now remembered by few. It is time to revive his analyses and criticism of the dominant testing culture. The timing is right for this revival since objective research is finally revealing the shortcomings of a ranking/grading culture and assessment methods – such as true-false and multiple-choice tests – and the limitations of high-stakes tests in predicting outcomes in schools and in the real world. Hoffman's work finally has the backing of statistics to demonstrate just how far ahead of the times he was.

Hoffmann does not argue against assessment in general, but he does argue vehemently against the dominant highly ritualized, well-oiled machinery that developed since the bureaucratization of education and the associated need to seek efficiencies that were developed in the latter part of the 19th century and early part of the 20th century, but at the expense of effectiveness and the rights of individuals. Let us analyze how he quite rightly "rocks the boat."

> Testing is no game. It is in deadly earnest. If tests are misused, the consequences can be far from trifling. Lives can be warped and careers ruined – as much by unwarranted promotions as by misguided guidance and lack of recognition of ability. The zest and creativity of a business organization may be dampened and destroyed. The strength and vitality of a nation may be jeopardized.

---

[4] A deceptive goal or hope. https://www.merriam-webster.com/dictionary/ignis%20fatuus

[5] Hoffman's is a very impressive life and his writings demonstrate a scholar of the first order. If only academia were made up of individuals like this. See https://en.wikipedia.org/wiki/Banesh_Hoffmann

> Tests are misused when they are taken too seriously. Though testing is no game, people in positions of responsibility would do well to treat it as one. Otherwise professional judgment becomes overawed and atrophied, and professional testers take over. (p. 103)

> Why should there be a standard of "scholastic aptitude" when, for all we know, there is no such specific thing as "scholastic aptitude" in the first place? Is it not a numerical invention? If by "scholastic aptitude" we mean some combination of relatively superficial, numerically measurable, verbal and arithmetical traits that are correlated with professorial grades with all their crotchets, idiosyncrasies, and exceptions, are we not talking of a smoothed-out, carefully sandpapered, statistical monstrosity…. (p. 136)

In speaking of multiple-choice tests, Hoffmann said, "They have a pernicious effect on education and the recognition of merit." (p. 150)

> [T]esting in school and college does the very opposite of what was hoped. In the one case the method represses individuality; in the other it misreads performance.
>
> …The desire to teach great numbers does raise difficulties correspondingly great. But it is no solution to do something next door to what is wanted simply because that something is easier to do. … Though this is not precisely the analogue of what we have done in the matter of examining the young learner's knowledge, it is precisely true of the arguments used in support of mechanical testing: it is easier and cheaper than the method of confronting mind with mind through the written word.
>
> The further argument that essay examinations cannot be graded uniformly, even by the same reader, only shows again the character of mind itself: it **is not an object to be weighed or sampled by volume like a peck of potatoes or a cord of wood**. Variations in performance and estimate will always subsist. Hence **an objective test of mind is a contradiction in terms**…. Anyone who has suddenly doubted the spelling of a word which he was about to write correctly will recognize how easily doubt can work distraction upon thought. (Emphasis added)
>
> …But if the tendency of such tests is to denature or misrepresent knowledge, to discourage the right habits of the true student, and to discriminate against the original in favor of the routine mind, of what use are such tests to a nation that has from its beginnings set a high value on instruction and the search for truth? (pp. 9-11)

Given the weaknesses and flaws in a grading system, psychometricians felt that teachers' subjectivity must be removed from grading and that "objective" science would resolve the problem with well-structured true-false and multiple-choice questions. In the following, Hoffmann points out the flaws with "objective science":

> Not everything is either true or false. Not everything is either black or white. Not every question can be answered either yes or no. It is not for nothing that our language has its *ifs* and *buts*, its *yets* and *howevers*, its *nevertheless*es and *notwithstandings*, its *possiblies* and *probablies* and *perhapses*, and its *on-the-other-hands*. (p. 53)

> …The testers call their multiple-choice tests "objective tests" and would have us regard objectivity as a virtue. But the term "objective test" is a misnomer. The objectivity resides not in the test as a whole but merely in the fact that no subjective element enters the *process* of grading once the key is decided upon. (pp. 60-61)
>
> Any competent person who has ever graded a non-objective mathematics or science examination knows that a correct answer obtained by incorrect methods is worth very little, while a wrong answer obtained by correct methods can deserve a top score; and even that a wrong answer obtained by wrong methods can be indicative of outstanding ability, and merit a bonus score. For example, the problem as stated in the examination might be a fairly straightforward one, but the student might misread it as a much harder problem. This harder problem might actually be beyond the capabilities of even the best students at his level to solve. But if he nevertheless made a brilliant attempt, and if this attempt failed because of subtle reasons that he could not be expected to perceive, then the wise examiner who knew his subject well would realize that he was dealing with a student of unusual ability.
>
> Similarly, any competent person grading a non-objective examination in history, sociology, and the like, will know that an incorrect conclusion arrived at by excellent arguments must be regarded as having far greater merit than a correct conclusion arrived at by appalling illogic, and will assign his grades accordingly.
>
> But the professional objective testers ignore all this. They are concerned only with the final choice, not with the quality of the reasoning that led to it. They are prepared to make enormous sacrifices for the sake of achieving objectivity. If essay testers and interviewers and players of hunches were prepared to make equivalent sacrifices they too could achieve comparable objectivity, numerical nicety, and pseudo-scientific decorum. (p. 66)

Well, we now know that true-false and multiple-choice tests are actually highly subjective. So it appears "objective science" has failed us here. But we shouldn't be surprised. Science cannot provide all the answers even though many believe it should.

Perhaps we are missing the point and looking in the wrong direction. For one thing, is it appropriate to grade students in an attempt to rank them as superior or inferior in comparison to their peers? A second point: Is it appropriate for instruction and assessment to be focused on logical-mathematical and linguistic talents alone – as the system is currently structured?

If the other human talents were incorporated into the system, the academic types would probably cry foul if they were to be graded, judged, and ranked based upon the other human talents such as kinesthetic, spatial, musical, etc. If we were to incorporate all human talents into an educational program, with appropriate assessment strategies, we would probably discover that the **vast** majority of students' grades would be a solid C, with very few achieving a B and none achieving an A. This is because no one possesses all talents at very high levels and when averaged out, we would discover that everyone is deficient in various areas and strong in others. So even though I may be very strong linguistically, I may be very weak spatially, while another individual's talents are the reverse of mine. When all talents are measured equally, this would average out as a C for me and everyone else.

This is not a human frailty as the assessment ranking system implies. In fact, it is a human strength. A community made up of diverse talents provides various abilities to contribute to the success of the community, as long as the various talents are sufficiently respected. This requires the abandonment of the status and prestige mindset. Therefore, the idea of ranking students based on two human talents – that is linguistic and mathematical-logical – at the exclusion of all others, is a gross injustice and extremely harmful to individuals and society. To put it bluntly, it's downright stupid; yet many academics genuflect at this altar, which reflects the degradation of scholarship and wisdom in the academic community.

The appropriate time to judge or rank students according to their abilities is when they choose a career direction in a given field, since they are attempting to become masters in it. After all, this is when it really matters since a community does not want professionals who are incompetent at their jobs. Until that time, pass/fail competencies are all that really matter since extensive mastery of any academic subject during the public-school years is irrelevant and not possible anyway. If a student has proven competent in a given requirement, a passing grade, without ranking, is all that is required. In this way we avoid bloating egos of some while avoiding the destruction of them in others. We need to abandon the winner-loser mindset in our public education system. There is enough of this evil in the real world. We don't need to elevate it as a virtue through our public institutions. In fact, these institutions should be promoting charity, compassion, camaraderie, solidarity, harmony, cohesion, commonality, etc., throughout our society. The primitive win-lose mindset should be left to the private sector where people have the freedom of association to knock one another around in an effort to be the knight in shining armor who vanquishes his opponent in the joust. Those who do not wish to attend the tournament should not be forced to submit to its rules through government coercion.

What would be the harm for most, or even all, test-takers to get excellent grades on tests? What is the need for "fooling" test-takers by traps? It becomes obvious that the purpose is to pick "winners" from "losers" – i.e. to eugenically select. This begs the question, "Why?" In a society based on freedom and equality, isn't this the antithesis of our founding perspective?

Perhaps it is to achieve cultural indoctrination? In other words, the Progressive plan was for academia to produce the leaders of society who had been thoroughly inculcated with the "correct" political outlook; and those who could not pick the "best" answers were people who could not easily be manipulated. Hoffmann alludes to this when he says:

> It is obvious from the nature of the tests that [test-makers] do not give the candidate a significant opportunity to express himself. If he is subtle in his choice of answers it will go against him; and yet there is no other way for him to show any individuality. If he is strong-minded, non-conformist, unusual, original, or creative – as so many of the truly important people are – he must stifle his impulses and conform as best he can to the norms that the multiple-choice testers set up in their unimaginative, scientific way. The more profoundly gifted the candidate is, the more his resentment will rise against the mental strait jacket into which the testers would force his mind. And if, by the questions they use, the testers betray intellectual incompetence, the profound student can hardly escape a feeling of contempt – contempt tinged with dismay that these are the people who have acquired the power to judge him.

> As for motivation, what chance has the candidate to show even that he is capable of sustained, probing mental effort when the tests skitter breathlessly from question to question? These tests favor the nimble-witted, quick-reading candidates who form fast superficial judgments. Some of these high-scoring candidates are extraordinarily able, of course; they are the ones who happen to have also at least some of the important attributes that the tests fail to detect. But other high-scoring candidates are meretricious and lack intellectual substance; yet they outscore their betters. (pp. 91-92)

Hoffmann points out the ambiguities that exist in college assessment tests and how they seek to trick test-takers as though seeing through their subjective and arbitrary tricks reveals "intelligence."

> [I]n a booklet *Scholastic Aptitude Test*, published by the College Entrance Examination Board in 1956, describing tests given to students seeking admission to college, and giving sample questions, the following advice is offered on page 18:
>
> > As you read through the explanations of the verbal section, you may disagree with what we think to be the correct answer to one or two questions. You may think we are quibbling in making certain distinctions between answer choices. It is true that you will find some close distinctions and just as true that in making close distinctions reasonable people do disagree. Whether or not you disagree on a few questions is not terribly important, however, for the value of the test as a whole is that people who are likely to succeed in college agree in the main on most of the correct answers. It is this that gives the [SAT] its predictive power. (p. 74)

In other words, only like-minded people need apply. It predicts if you fit into the academic culture rather than if you truly have scholarly abilities. This is selection based on subjective values – perhaps, more eugenics in play.

This is a very good example why businessmen have very little faith in many college graduates. Academia is a fantasy world that breeds arrogance in many who pass through its system. Such people can be worse than useless in a business setting since their pompousness makes them intolerable to work with, and their lack of real-world abilities means they will make very poor decisions and expect others to follow them without question. That's a recipe for disaster for business.

Hoffman offers a very nice example of the weakness of multiple-choice tests. The following example is exactly what I experienced throughout my years in the system and it is a major contributor to my extremely critical view of the educational establishment.

> I am told that on a certain test a question appeared of which the following is the gist:
>
> *Emperor* is the name of
>
> (A) a string quartet
> (B) a piano concerto
> (C) a violin sonata

> This seems to be a simple, straightforward question. The average student quickly picks answer B and proceeds untroubled to the next question, perhaps feeling elated at being given so simple a test. But what of the superior student? He knows of the *Emperor Concerto* of Beethoven, but he also knows of the *Emperor Quartet* of Haydn; and his knowledge puts him at a disadvantage, for because of it he must pause to weigh the relative merits of answers A and B while his more fortunate, less well-informed competitors rush ahead.
>
> In this particular case the superior student does not ponder long. Two theories occur to him: the examiner is malicious,[6] or the examiner is ignorant of the Haydn work. If this is the first dubious item that he has encountered on the test, he inclines to the second alternative and chooses answer B with little delay.
>
> Yet even in this simple case he suffers a penalty far exceeding the slight loss of time. For he has been led to call into question both the good will and the competence of the examiner; and this subjects him to a psychological handicap, the severity of which will depend on how faulty or impeccable is the rest of the test. No longer is it possible for him to skim innocently ahead. Instead, he must proceed warily and dubiously, ever alert for intentional and unintentional pitfalls. And whenever he comes to a question for which he, with his superior ability, sees more than one reasonable answer, he must stop to evaluate afresh the degrees of malice and incompetence of the examiner. Such a test becomes for the superior student a highly subjective exercise in applied psychology – and, if he is sensitive, an agonizing one. (pp. 22-23)

Another point to be made of this example is the uselessness of the knowledge this test question reflects, unless one is a music major. While such information may be interesting, it does not pass Herbert Spencer's *relative worth* scale. That is: All knowledge has worth, but each bit of knowledge must be measured against the value of other knowledge that will be sacrificed since we cannot learn everything there is to know. A cost-benefit analysis is required for everything that is taught and then such an analysis is required to determine its relative worth as it relates to memory retention and exams. If it is simply nice to know, then there is no justification in including it in a test, and measuring memory and recall abilities are no justification either since they are not a measurement of intelligence.

The following email message that was circulating a few years ago demonstrates that what a test-maker believes he is asking isn't always the way the test-taker interprets it. While these are very funny and silly examples, they exemplify why such testing is troublesome.

"How would <u>you</u> score this exam? Each answer is absolutely grammatically correct, and funny too.

Q. 1. In which battle did Napoleon die?
A. His last battle.

---

[6] Think about this: Trickery in formulating questions by test-makers condones underhanded behavior, which is a terrible example for our children to discover what academia is all about. That is, in order to succeed in academia and have access to the credentials it provides, one must think in devious ways if one is to survive. It establishes institutional immorality at a very young age. If this is not rectified, it is better to close down such public institutions than to promote and perpetuate social immorality, especially since trickery such as this has no relation to intelligence.

Q. 2. Where was the Declaration of Independence signed?
A. At the bottom of the page.

Q. 3. River Ravi flows in which state?
A. Liquid.

Q. 4. What is the main reason for divorce?
A. Marriage.

Q. 5. What is the main reason for failure?
A. Exams.

Q. 6. What can you never eat for breakfast?
A. Lunch & dinner.

Q. 7. What looks like half an apple?
A. The other half.

Q. 8. If you throw a red stone into the blue sea, what will it become?
A. Wet.

Q. 9. How can a man go eight days without sleeping?
A. No problem, he sleeps at night.

Q. 10. How can you lift an elephant with one hand?
A. You will never find an elephant that has one hand.

Q. 11. If you had three apples and four oranges in one hand and four apples and three oranges in other hand, what would you have?
A. Very large hands.

Q. 12. If it took eight men ten hours to build a wall, how long would it take four men to build it?
A. No time at all, the wall is already built.

Q. 13. How can you drop a raw egg onto a concrete floor without cracking it?
A. Any way you want, concrete floors are very hard to crack."

For the test-maker, this would come as a shock to see such interpretations. Perhaps this is because a person whose mind has accepted a certain perspective, finds another perspective difficult to perceive. This is due to what Willingham (2007) points out: The human brain does not like to think; it prefers habituation and doing things automatically since it is designed this way. Therefore, test-makers' minds are habituated to the testing industry's perspective – with its own cultural idiosyncrasies – and if the test-taker is unfamiliar with this cultural perspective, poor performance on tests is to be expected, regardless of intelligence. This helps explain the argument that minorities are at a disadvantage when it comes to tests since they are not as familiar with the Caucasian test makers' culture. This is

inappropriately alleged to propagate “white privilege” but it is easy to understand how minds with narrow perspectives might come to this conclusion since it really is a dilemma that demands a remedy.

Another humorous email in circulation demonstrates how statements are not always perceived the way one might expect them to be interpreted:

SMART ASS ANSWER #1
It was mealtime during an airline flight. 'Would you like dinner?', the flight attendant asked John, seated in front.

'What are my choices?' John asked.

'Yes or no,' she replied.

SMART ASS ANSWER #2
A lady was picking through the frozen turkeys at the grocery store but she couldn't find one big enough for her family. She asked a stock boy, ' Do these turkeys get any bigger?'

The stock boy replied, 'No ma'am, they're dead...'

SMART ASS ANSWER #3
The police officer got out of his car as the kid who was stopped for speeding rolled down his window. 'I've been waiting for you all day,' the officer said.

The kid replied, Yeah, well I got here as fast as I could.

When the cop finally stopped laughing, he sent the kid on his way without a ticket.

SMART ASS ANSWER #4
A truck driver was driving along on the freeway and noticed a sign that read: Low Bridge Ahead. Before he knew it, the bridge was right in front of him and his truck got wedged under it. Cars are backed up for miles.

Finally a police car comes up. The cop gets out of his car and walks to the truck driver, puts his hands on his hips and says, 'Got stuck, huh?'

The truck driver says, 'No, I was delivering this bridge and I ran out of gas.'

This shows that we must be very careful in how we design assessment tests when so much is riding on the test takers performance. If we rely on interpretation, then the test may very well be worthless, but the consequences can be catastrophic to the future of individuals marginalized by the assessment regime.

* * *

Next Hoffmann provides an excellent example where the test-taker is expected to pick the “best” answer. He sets up a somewhat lengthy scenario that I will summarize for the reader’s convenience. The

question posed is multiple choice, which requires a pair of words to be chosen that fills two different blanks in a sentence. The problem is, none of the combinations are correct.

> Unfortunately, this does not release you from the requirement of making a specific choice. The rules of this particular type of multiple-choice question call on you to pick the "best" answer, even though having to do so is a hardship when no answer is even good.
>
> How do you go about picking the "best" answer here? Perhaps you reject [one] answer on purely grammatical grounds, since [it] does not go well with [its partner] – although you could not safely do even this had there been a grammatical error in the directions. You then study the other answers carefully, **trying to decide which one is least objectionable**. (Emphasis added)
>
> Suppose in the end you realize that [another] answer is [the right] one. You still cannot help wondering whether [that] answer was the one you were really supposed to pick, for you find yourself involved in a subjective guessing game, wondering not which answer is intrinsically the "best" but, instead, which answer the poser of the question may happen to think is "best." (pp. 23-25)
>
> What sense is there in giving tests in which the candidate just picks answers without any opportunity to give reasons for his choices? (p. 29)
>
> [T]here is no satisfactory method of testing – nor is there likely to be. Human abilities and potentialities are too complex, too diverse, and too intricately interactive to be measured satisfactorily by present techniques. There is reason to doubt even that they can be meaningfully measured at all in numerical terms. (p. 30)

Picking the "best" answer because it reflects the likelihood of success in academia shows it is indeed a form of eugenics selection since, after many generations, the academic community will approach clone-like characteristics, with individuality and creativity becoming recessive traits.

Hoffmann then states:

> Rarely do the multiple-choice testers ask the candidate to pick the "correct" answer. … Complain to them, for example, that the wanted answer to a particular question is certainly not a correct one and is not really even a good one, and they will point out that all is well since it is nevertheless "the best." (p. 77)

Requiring test-takers answer with the "best" answers rather than the right answers does not demonstrate an intelligent person. It demonstrates the ability to see things in a similar manner as test-makers – which is a contributing factor to minority challenges with such tests, since culture plays a part in test-makers' perspectives.

It is obvious the system is weak in rewarding greater abilities due to the very narrow and limited minds of test-makers whose questions and answers tend to reward mediocrity, which is the natural consequence of centralized bureaucratic systems.

When we consider the malice of test-makers, the mistakes they make in formulating tests, the irrelevance of many questions and answers, one wonders what these tests are designed to accomplish. Such idiocy in the real world does not endure for long, so why judge students, and play god with their futures, based on such absurdity?

Next, Hoffman addresses the problem with the "halo effect" and subjectivity in those who grade essay papers. Regarding the "halo effect," he makes reference to "secondary matters little related to the content and style of the essay" such as handwriting, spelling, punctuation, etc., that may lower the grader's opinion of the writer's abilities, thereby effecting the way a paper is graded. "A rating of a single skill is not a self-contained thing. It carries a halo that colors other ratings." (p. 46)

Regarding subjectivity, Hoffmann reveals the problem test psychologists had discovered about subjective judgments of graders. "[Test psychologists] take, for example, a single essay and ask various qualified people to grade it. The discrepancies in the grades are frightening. The test psychologists tell us that a single essay can receive all grades from A to F when graded independently by different graders." (p. 46)

As for the length of tests, Hoffmann informs us that test-makers brag about how extensive their tests are. "[T]hey speak glowingly of the great efficiency of their tests and the enormous amount of material the tests can cover in a short time." However, Hoffmann warns against being lured into this trap.

> Breadth and efficiency can easily be made to seem admirable. But they are not, in themselves, necessarily desirable, and they can be dearly bought when the price is depth and scope for creativity. Broad coverage without depth can favor the candidate who has superficial knowledge of many things and profound knowledge of nothing. The people who know the most are not always the most valuable people. (p. 87)

Hoffmann then provides an anecdote of a Nobel prize winner who pointed to colleagues who might have been more cleaver than he, but who were unable to focus and finish their work. He then quoted Genghis Khan who said: "it is the completion which gives value to an action." Cleverness might be "cute" but is useless if unable to be applied.

Next, Hoffmann quotes a psychologist, John Shlien, who wrote the article "Mental Testing and Modern Society" (*The Humanist*, 1958):

> Tests are sold to us on the basis of their "efficiency." Aside from the validity of this claim, the idea of efficiency itself needs to be re-examined. It is often a short-term concept, and a short-sighted one. Suddenly speaking, the most efficient way to get exactly the proteins you need is to take a bite of the nearest person. To get to the ground floor, jump out the window. But these very immediate goals are not our complete or real ones. Until we have thought about these, we cannot use efficiency, even if it can be delivered. Long-term efficiency may rest much more upon people going where and doing what they want than on placing them where tests say they fit best. (p. 90)

Hoffmann points out that the test-makers "had to find out why the tests were sometimes so strikingly fallible; and they discovered, apparently not without surprise, that the tests did not measure motivation, creativity, and other important ingredients of greatness." (p. 94) Then he offers the following:

> It is not the presence of defective questions that makes multiple-choice tests bad. Such questions merely make them worse. Even if all questions were impeccable, the deep student would still be at a disadvantage. He would see more in a question than his more superficial competitors would ever dream was in it, and would expend more time and mental energy than they in answering it. That is the way his mind works. That is, indeed, his special merit. But the multiple-choice tests are concerned solely with the candidate's choice of answer, and not with his reasons for his choice. Thus they ignore that elusive yet crucial thing we call quality. (p. 99)

Hoffmann (p. 41) points out how powerful the testing industry has become and the dangerous tendencies it entails. Given the current destructive system, we know how frustrated many teachers become and why many simply quit teaching since hurtles to improvement have grown to the size of insurmountable mountains.

It is also understood that we lose a lot of students who would otherwise have respect for the educational system. When they realize that the system is trying to marginalize them through trickery – amongst other strategies –, so "winners" can be winnowed from the "losers," bitterness sets in, especially for those who know they lack the talents that the academics and bureaucrats have optimized the system for. Many students then realize this is a game they cannot win; hence they stop playing to whatever degree that suits them – be it quitting school or simply doing the bare minimum to get through "prison." The educational establishment then writes them off and feels good that it has produced "winners" – i.e. the genetically "superior" portion of society – who will go on to represent it in a "positive light" by going to college. The "losers" will then have to learn how to navigate on their own through an extremely complex society that has pitfalls and snares dotting the economic and legal landscape. Ambitious politicians then prey on this large sector with ideas of "hope and change" which – regardless of intent – equates to a steady march toward totalitarianism.

The testing industry, combined with the accreditation and credentialing system, can, for all intents and purposes, be compared to the "military industrial complex" that is, a leviathan with unique powers and dangers to the country. It is a decadent system that harms children to further its own interests. Better to return to a time when most children stopped school somewhere between 6th and 8th grade rather than continue such destructive institutional forces. In this way, parents could help their children discover other paths to pursue during their critical formative years. Anyone with even a marginal semblance of wisdom who is made aware of this destructive institution, would realize that a privatized system or locally run system, though not perfect, would be **far** better than the current highly centralized government institutionalized system we have.

To conclude this section, I think Guskey (2015) delivers the final nail in the coffin proving how tyrannical the testing regime can be. He shows the methods used by many in the academic community to sort students into groups of winners and losers. He discusses how it is done intentionally and maliciously. It then becomes apparent that this is used to create a synthetic aristocracy of which they are a part.

The Antithesis of Applied Education

Lave (1988) discusses the *functionalist sociology of education* theory[7]:

> Functional theory underlies the web of relations between academic, novice and just plain folk "worlds." … But it may not be as well understood that the functionalist position contains a theory of learning: in particular, that children can be taught general cognitive skills (e.g. reading, writing, mathematics, logic, critical thinking) if these "skills" are disembedded from the routine contexts of their use. Extraction of knowledge from the particulars of experience, of activity from its context, is the condition for making knowledge available for general application in all situations. Schooling reflects these ideas at a broad organizational level, as it separates children from the contexts of their own and their families' daily lives. At a more specific level, classroom tests put the principle to work: they serve as the measure of individual, "out of context" success, for the test-taker must rely on memory alone and may not use books, classmates, or other resources for information. Arguably examinations are also condensed, symbolic, ritual ordeals which inculate [sic] the essence of the theory.
>
> Cognitive psychology accounts for stability and continuity of cognitive activity across settings through the psychological mechanism of learning transfer. That is, knowledge acquired in "context-free" circumstances is supposed to be available for general application in all contexts, widely transportable but relatively impervious to change in the course, and by the process, of travel and use.[8] (pp.8-9)
>
> Functionalists argue that verbally transmitted, explicit, general knowledge is the main prerequisite that makes cognitive skills available for transfer across situations. Social practice proponents argue that knowledge-in-practice, constituted in the settings of practice, is the locus of the most powerful knowledgeability of people in the lived-in world. In the functionalist view the label "everyday" is heavy with negative connotations emanating from its definition in contrast to scientific thought. (p. 14)
>
> It is puzzling that learning transfer [theory] has lasted for so long as a key conceptual bridge without critical challenge. The lack of stable, robust results in learning transfer experiments as well as accumulating evidence from cross-situational research on everyday practice, raises a number of questions about the assumptions on which transfer theory is based – the nature of cognitive "skills," the "contexts" of problem-solving and "out of context" learning, the normative sources of models of good thinking and less than perfect "performances." Transfer

---

[7] "The functionalist theory focuses on the ways that universal education serves the needs of society. Functionalists first see education in its manifest role: conveying basic knowledge and skills to the next generation. Durkheim (the founder of functionalist theory) identified the latent role of education as one of socializing people into society's mainstream. This 'moral education,' as he called it, helped form a more-cohesive social structure by bringing together people from diverse backgrounds, which echoes the historical concern of 'Americanizing' immigrants. "Functionalists point to other latent roles of education such as transmission of core values and social control. The core values in American education reflect those characteristics that support the political and economic systems that originally fueled education. Therefore, children in America receive rewards for following schedules, following directions, meeting deadlines, and obeying authority. …" https://www.cliffsnotes.com/study-guides/sociology/education/theories-of-education

[8] Of course, this is the *functionalist theory*. This does not define cognitive psychology.

> theory may well owe its longevity to its central location in the web of relations … institutionalized in divisions between the disciplines of anthropology and psychology, in schooling, and in dichotomies between scientific and everyday thought. Basic and profoundly embedded assumptions govern the persistent loyalty to transfer [theory] and all that it stands for, and a strong break with this tradition, though costly in theoretical consensus, is a promising means for moving the study of cognition into the larger social world (p. 19).

Since Lave's excellent work in dispelling the functionalist transfer theory, many cognitive psychologists have fully demonstrated that the theory of "out of context learning" – i.e. non-applied methods – to achieve transfer as being ineffective, but it is such an integral part of our educational system – being deeply ingrained in the subconscious of many educators – that it has shown itself to be extremely difficult to dislodge. This helps explain why what is taught, how it is taught, and how poorly our assessment methodologies are designed has proven to be so difficult to change since the change would require the academic community's underlying assumptions to be, in large part, abandoned.

The Current State of U.S. Education and Skills

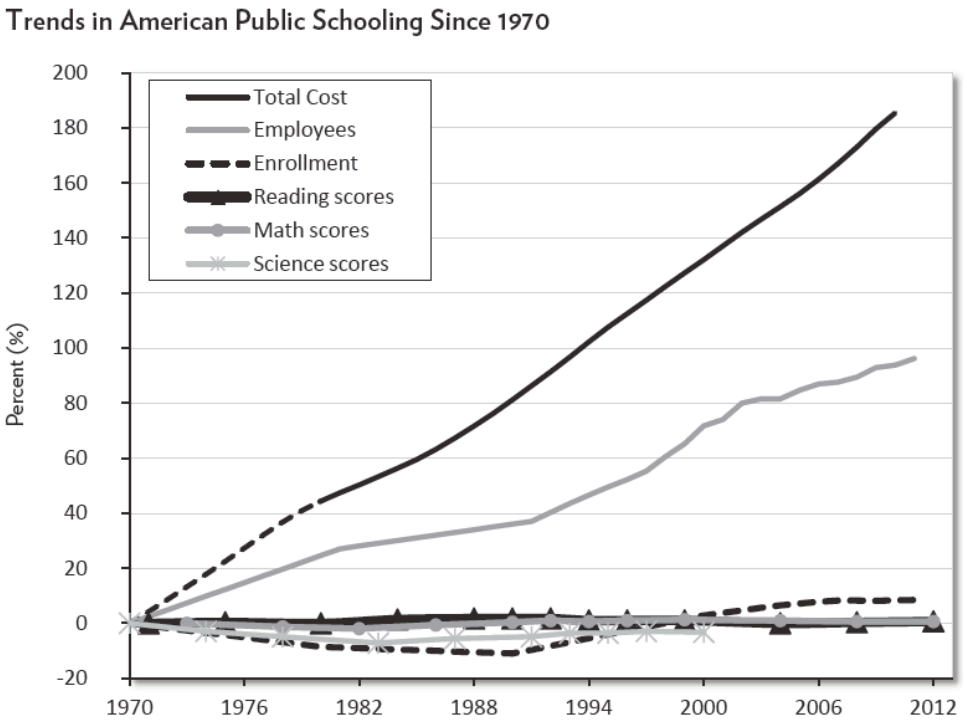
The Educational Testing Services (Goodman et al., 2015) offers a report that paints a grim picture of the current state of affairs for the United States. This report focused on 16- 34-year-olds at the time of the assessment – i.e. "millennials." They utilized another report published by the Organization for Economic Cooperation and Development (OECD) titled *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*. The report compares the skills-performance in literacy, numeracy, and problem solving … of 16- to 65-year-old U.S. adults to other nations. The survey used to gather the data is titled *Programme for the International Assessment of Adult Competencies* (PIAAC). "The report revealed that the skill levels of U.S. adults compared to those of 21 other participating OECD countries were dismal across the board." (p. 6)

> **One central message that emerges from this report is that, despite having the highest levels of educational attainment of any previous American generation, these young adults on average demonstrate relatively weak skills in literacy, numeracy, and problem solving** … compared to their international peers. These findings hold true when looking at millennials overall, our best performing and most educated, those who are native born, and those from the highest socioeconomic background. **Equally troubling is that these findings represent a decrease in literacy and numeracy skills for U.S. adults when compared with results from previous adult surveys**.[9] (p. 2) (Emphasis added)
>
> The report points to the challenges that have evolved over the last 40 years and how the political and economic landscape has shifted, benefiting those with more education and how those with less are falling further behind. This trend is not expected to improve. [S]hifts in the nature of work and demands for skilled vs. unskilled labor have made it even more apparent today that individual's economic security and prosperity rest in large measure on the acquisition of specific skills as well as the ability to augment skill proficiency throughout one's lifetime. (p. 7)

[9] This demonstrates that even though this cohort has more education at a much greater expense than previous generations, it provides poorer quality.

> … As the authors note, these changes have both immediate and long-term consequences for families, communities, and the nation as a whole. **The findings also offer a clear caution to anyone who believes that our policies around education should focus primarily on years of schooling or trusts that the conferring of credentials and certificates alone is enough**. While it is true that, on average, the more years of schooling one completes, the more skills one acquires, this report suggests that far too many are graduating high school and completing postsecondary educational programs without receiving adequate skills. If we expect to have a better educated population and a more competitive workforce, policy makers and other stakeholders will need to shift the conversation from one of educational attainment to one that acknowledges the growing importance of skills and examines these more critically. How are skills distributed in the population and how do they relate to important social and economic outcomes? How can we ensure that students earning a high school diploma and a postsecondary degree acquire the necessary skills to fully participate in our society? (p. 2) (Emphasis added)

More education and more money being spent on education has proven to be a failing strategy given the current educational structure and goals. The evidence is overwhelming; the U.S. is losing ground. Just consider how the increases in educational spending coupled with increases in educational administrators have had no effect on educational outcomes for students. Quality has been sacrificed for quantity proving the current educational trajectory is a complete failure. Only bureaucratic systems can ignore such obvious results.



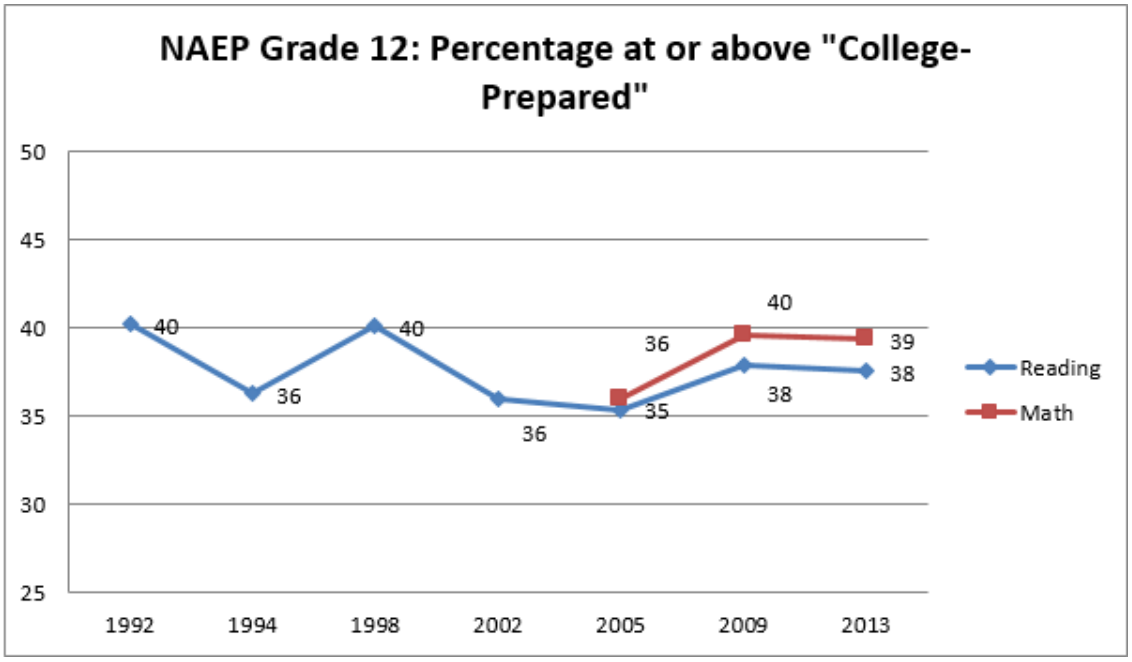
Sources: U.S. Department of Education, "Digest of Education Statistics"; and NAEP tests, "Long Term Trends, 17-Year-Olds."
Note: "Total cost" is the full amount spent on the K-12 education of a student graduating in the given year, adjusted for inflation. In 1970, the amount was $56,903; in 2010, the amount was $164,426.

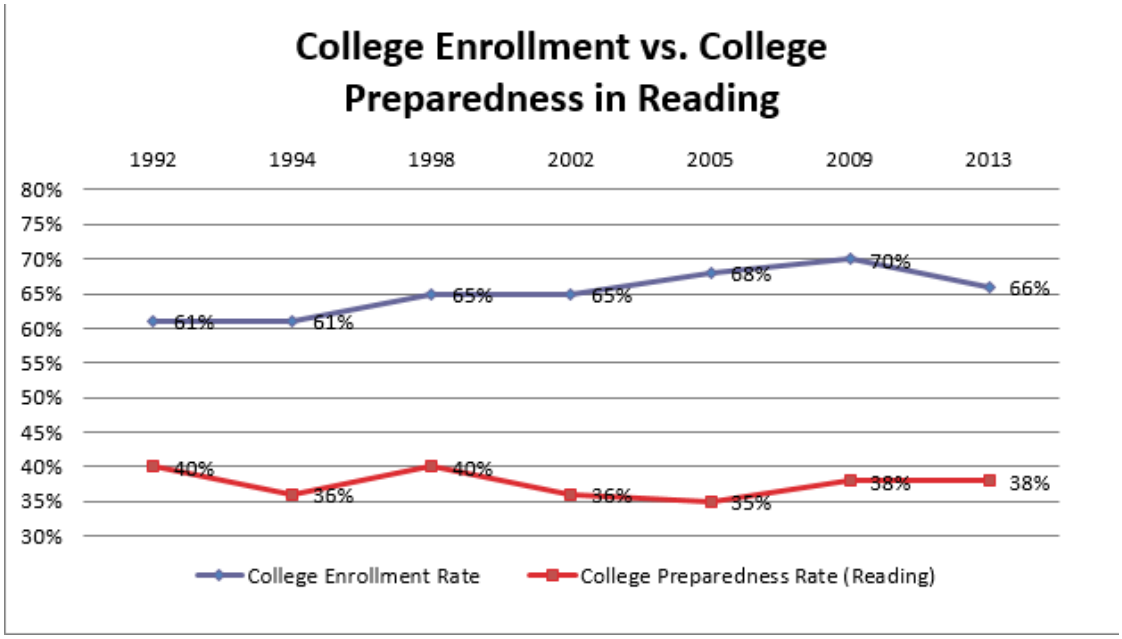
State Education Trends: Academic Performance and Spending Over the Past 40 Years.

Cato Policy Analysis, Mar. 18, 2014, Number 746.

The report mentions the decline in adult skills, expressing real concern for this trend.

> The average score for U.S. adults in literacy has declined since 1994. In numeracy, average scores for adults have declined since 2003. … If we continue on this path, there could be serious consequences for America's economy and the future prosperity of our workforce.





College Preparedness Over the Years, According to NAEP, Fordham Institute, 2015

> The individual and societal costs of having a large proportion of the population with low skills (both compared to previous years and to the percentages in other countries for the PIAAC assessment) should not be underestimated. In addition to economic costs in terms of the competitiveness of the labor force in a global economy, there are also more subtle, no-less-important consequences of having a population with deeply divided skill levels. Such societies risk becoming increasingly polarized, fragmented, and divided. Social cohesion suffers, and civic engagement becomes more sporadic and tenuous. Cross-country comparisons of the PIAAC results by OECD reveal some important and clear patterns and correlations: Adults with higher skills are more likely to report better health, have more trust in political institutions, and demonstrate increased rates of volunteerism. Lower skilled adults, on the other hand, represent the most vulnerable members of society – those at risk of having restricted access to basic services and less than full participation in democratic practices and educational opportunities. Skills are also strongly associated with access to labor participation and training opportunities. Other research suggests that the distribution of skills of a country's population is inextricably – albeit complicatedly – linked to the distribution of its income and wealth. (p. 8)
>
> … The ripple effects can also be felt in terms of increased income and wealth inequality. In fact, the entire society is affected by a cycle that perpetuates and exacerbates inequalities and brings into question whether we are offering individuals an equal opportunity to succeed. As the economist Joseph Stiglitz has pointedly acknowledged, an economic and political system that is perceived to favor some citizens over others is not sustainable in the long run. "Eventually," Stiglitz warns, "faith in democracy and the market economy will erode, and the legitimacy of existing institutions and arrangements will be called into question." Although the nature of the relationship among the distribution of skills, economic inequality, and challenges to sustaining a coherent participatory democracy is part of a complex story with many narrative threads, it is an important one to explore and understand. (p. 9)

I believe the Educational Testing Service has done a tremendous service to the U.S. by offering this extremely insightful and forward-looking report that must be seen as a clarion call to action. They leave no doubt of the dire consequences of continuing down the same worn-out failing path we've been traveling. They are an organization that Applied Education Foundation seems to have much in common with and therefore should seek collaboration.

I would like to offer a prediction here. With high unemployment levels having dominated the U.S. for the last decade (I write this in January 2017), and with President-elect Donald Trump yet to take office, I predict that a Republican dominated government (both houses and the presidency) will usher in unprecedented growth which will create significant pressures on the labor market, thereby driving wages on an upward trend. Reports, such as the Educational Testing Service's will, in the future, be looked upon as a gloom and doom report and that they got it wrong. I caution everyone to avoid this optimistic view while the economy is booming. Hard times will come again since economic cycles are a reality of civil life. Therefore, I urge the country to invest heavily in reengineering education while times are good so that when the economy takes a fall again, it won't be a long drop and it won't be as painful as it would otherwise be. Let's learn from history rather than see it repeated over and again.

Schwartz et al. (2005) offer the following:

> … Broudy has attempted to understand the kinds of educational experiences that prepare students for life rather than simply for test taking:
>
> "Ever since formal schooling was established, it has been assumed that knowledge acquired in school would be used to enhance the quality of human life. The investment in schools was supposed to yield a return in the form of greater adequacy in occupational, civic, and personal development."
>
> Broudy discusses the "replicative," "applicative," and "interpretive" aspects of knowing and notes that most assessments have focused almost exclusively on the first two. For example, he argues that students rapidly forget the facts that they learn in school, hence they do poorly as measured by tests of "replicative" knowing. He also argues that most students have difficulty applying previously acquired knowledge to solve new (transfer) problems; hence they do poorly in what he calls "applicative knowing." In fact, Broudy emphasizes that a sole reliance on replicative and applicative tests of knowing lead to the conclusion that schooling has very disappointing effects on lifelong learning. Replicative and applicative tests make school "look dumb"…. (pp. 8-9)

Intelligence Quotient (IQ)

[IQ testers] stand harsh guard at the gateway to National Merit Scholarships, and they tell admissions officers how many points' worth of college aptitude we possess. Hoffmann, p. 20, 1964/2003

In defining the word intelligence, it appears test-makers believed they had super human powers. Hoffmann's analysis is really quite revealing.

> Originally, IQ tests were indeed called *intelligence tests*, but this specialized use of the loose, popular word "intelligence" proved confusing. People who considered themselves intelligent thought they knew, at least approximately, what they meant when they used the word "intelligence" intelligently, and they wondered whether the intelligence testers were measuring the sort of thing the word "intelligence" suggested. There was considerable technical discussion of the point among psychologists, and ultimately intelligence testers took refuge in the above operational definition that *intelligence* is that which is measured by means of intelligence tests. (p. 108)

To think, the creation and design of tests is what creates intelligence. Test-makers must have divine revelation to be capable of this feat.

The history of the IQ test provides some insight into what many believe "intelligence" is. Its misuse in determining children's future is a perfect example of why a centralized system of control is utterly incompatible with liberty. It has literally determined the fate of millions of Americans over several generations. This is a travesty of the first order! And it will be recorded in history as a major failure of the period in which the U.S. embraced and applied Statist principles (called Progressivism in the U.S.) – i.e. the religion that marries centralized government with science to bring about utopia.

The assessment testing system we know today came in vogue at the turn of the last century through the American Progressive *mental measurement movement*. Kliebard (2004) points out how it was closely tied to the proponents of the Progressive *social efficiency movement*. The mental measurement movement "provided the technology necessary for the kind of assessment and prediction that a curriculum based on social efficiency doctrine required." (p. 89) However, Kliebard does point out that not all educational reformers were loyal to the mental measurement movement. William C. Bagley of Teachers College "had … been one of the very few leaders in education to see significant antidemocratic tendencies in the mental measurement movement that swept the country…." (p. 191)

The tool the mental measurement proponents found that fit their scheme was a test developed by Alfred Binet of France. At the turn of the last century, Binet was charged with finding the means to identify French students who needed special education and could be separated from normal learners in order to avoid disruption of classes. This assessment system eventually came to be known as the Binet Scale. An essay provided by Edublox explains it this way:

> Binet himself cautioned against misuse of the scale or misunderstanding of its implications. According to Binet, the scale was designed with a single purpose in mind; it was to serve as a guide to identify children in the schools who required special education, its intention was not to be used as "a general device for ranking all pupils according to mental worth." Binet also noted that "the scale, properly speaking, does not permit the measure of intelligence, because intellectual qualities are not superposable[10], and therefore cannot be measured as linear surfaces are measured." Since, according to Binet, intelligence could not be described as a single score, the use of the Intelligence Quotient (IQ) as a definite statement of a child's intellectual capability would be a serious mistake. In addition, Binet feared that IQ measurement would be used to condemn a child to a permanent "condition" of stupidity, thereby negatively affecting his education and livelihood….

However, to Binet's dismay, American Progressives perverted the scale's purpose.[11] Gould (1996) explains, "In the hands of [American] psychologists such as H. H. Goodard, Lewis H. Terman, R. M. Yerkes, and Edward L. Thorndike, that scale became not just a diagnostic device, but a powerful tool by which society could be regulated."

> H.H. Goddard … translated Binet's work into English and advocated a more general application of the [test]. Unlike Binet, Goddard considered intelligence a solitary, fixed and inborn entity that could be measured.
>
> … Lewis M. Terman, who also believed that intelligence was hereditary and fixed, worked on revising the [test]. His final product, published in 1916 as the *Stanford Revision of the Binet-Simon Scale of Intelligence* (also known as the *Stanford-Binet*), became the standard intelligence test in the United States for the next several decades. Once American educators had been convinced of the need for universal intelligence testing, and the efficiency it could contribute to

---

[10] To lay (something, such as a geometric figure) upon another so as to make all like parts coincide: superimpose.

[11] See the Appendix at the end of this essay, which provides extensive notes from Gould's book, *The Mismeasure of Man* (1996), to get a more thorough understanding of how this unfolded and how it affected the trajectory of American assessment testing.

> school programming, within a few years … was transformed into an integral, far-reaching component of the American educational structure. Through Goddard's and Terman's efforts the notion that intelligence tests were accurate, scientific, and valuable tools for bringing efficiency to the schools resulted in assigning the IQ score an almost exalted position as a primary, definitive, and permanent representation of the quality of an individual. Hence, intelligence testing became entrenched in the schools over the next several decades.
>
> Few people realize that the tests being used today – of which the IQ test continues to be the most popular – represent the end result of a historical process that has its origins in [intellectual] bigotry. Many of the founding fathers of the modern testing industry –including Goddard, Terman and Carl Brigham (the developer of the *Scholastic Aptitude Test*) – advocated eugenics. Eugenics is a movement concerned with the selective breeding of human beings. Selected human beings would be mated with each other in an attempt to obtain certain traits in their offspring, much the same way that animal breeders work with champion stock. The eventual goal of eugenics is to create a better human race. The Nazis took this idea to the extreme. All "inferior" humans, retarded children or adults, and any individuals with genetic defects, were to be destroyed….
>
> The founding fathers of the testing industry saw testing as one way of achieving the eugenicist aims. Goddard's belief in the innateness and inalterability of intelligence levels, for example, was so firm that he argued for the reconstruction of society along the lines dictated by IQ scores…. (Edublox)[12]

As this shows, Binet's Scale was transformed into purposes that served Progressivism, which was to identify "winners" and separate them from "losers" and to assign everyone to their appropriate station in the new world order. The test was appointed the name Intelligence Quotient (IQ) which was to measure "general intelligence" (which was assigned the symbol "*g*") of U.S. citizens. IQ was to determine which social stations individuals were capable of filling. For those with IQs deemed so low as to be defined by eugenicists as "imbeciles" or "morons," sterilization was to be performed on them so they could not reproduce. It is reported that Progressives sterilized 60,000 U.S. citizens between the 1920s and 1930s.

Lemann (1995) appropriately titled his article *The Great Sorting*, with the subtitle *The first mass administrations of a scholastic-aptitude test led with surprising speed to the idea that the nation's leaders would be the people who did well on tests*. This helps explain why so much effort is placed in trying to force everyone to do well on assessment tests. That is, everyone must be given the opportunity to become the nation's leaders, though at the expense of everyone else (if this sounds illogical, it is).

Lemann then goes into how things developed.

> In the April 1948 issue of a publication called *The Scientific Monthly* an article appeared under the title "The Measurement of Mental systems (Can Intelligence Be Measured?)." … The authors believed that intelligence tests were a fraud – a way of wrapping the fortunate children … in a mantle of scientifically demonstrated superiority.

[12] https://edubloxtutor.com/history-iq-test/.

Lemann then references the psychometricians that strongly influenced the direction IQ testing took. One of these important figures was Carl Brigham of the Educational Testing Service who was the father of the SAT. Such Progressives as this, and many others, believed that the IQ test proved that immigrants preceding WWI were intellectually inferior compared to certain Western races. IQs were also used to determine reproductive rights; that is, it was used to discourage reproduction and to sterilize some people.

However, there were those who did not accept the findings of such Progressives and offered a critique.

> Its three main tenets were: first, that the tests were measuring cultural conditioning rather than a biological trait; second, that there was no such thing as *g* – a single human ability more important than all others – but instead a group of human abilities; and third, that IQ tests could be misused to classify millions of young people as mentally inferior and so to deny them opportunity.

Of course, this is exactly what has happened throughout much of the 20th century and is still part of the education system.

Assessment tests provided institutional and social control rather than individual discovery and development. It is supposed to be a public good – that is, an educational system using public money – but has turned into a social evil where the majority of the population is marginalized through an educational system designed for the benefit of a minority of the population. With credentials being monopolized by this public system, the vast majority of citizens are denied access to economic well-being and full participation in the political realm.

Kaufman (2009) analyzes the relationship between real-life accomplishments and IQs:

> Average adult IQs associated with real-life accomplishments:
>
> - Neurosurgeons, research scientists, university professors 135+
> - MDs or PhDs 125
> - College graduates 115
> - 1–3 years of college 105-110
> - Clerical and sales workers 100-105
> - High school graduates, skilled workers (e.g., electricians, cabinetmakers) 100
> - 1–3 years of high school (completed 9–11 years of school) 95
> - Semi-skilled workers (e.g., truck drivers, factory workers) 90-95
> - Elementary school graduates (completed eighth grade) 90
> - Elementary school dropouts (completed 0–7 years of school) 80-85
> - Have 50/50 chance of reaching high school 75
>
> Average IQ of various occupational groups:
>
> - Professional and technical 112
> - Managers and administrators 104
> - Clerical workers; sales workers; skilled workers, craftsmen, and foremen 101

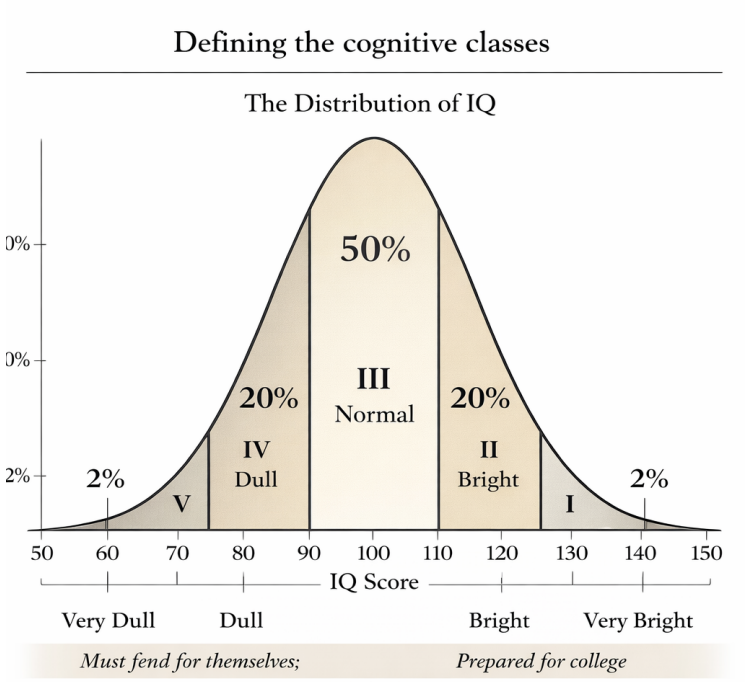
- Semi-skilled workers (operatives, service workers, including private household; farmers and farm managers) 92
- Unskilled workers 87

At first glance, one might think that "intelligence" measured by the IQ test has a direct correlation to abilities in a profession. However, if we delve below the surface of such vague data, we might begin to see something very different from our first impressions. First of all, the only way to the lucrative professions is through assessment tests, which have been demonstrated to measure memory and recall abilities for close transfer, but not actual real-life abilities and far transfer. So those without excellent memory and recall abilities will be prohibited from pursuing many professions even though they may have been outstanding at them. Perhaps some of these vocations require good memory and recall ability, but since poor test-takers are prohibited from entering many professions, we cannot know how they would have performed. In addition, it has been demonstrated that what IQ tests actually assess is what one learned rather than what one's potential is.[13] Also, the fortunes and misfortunes of life are known to play a huge role in IQ test results. Furthermore, for IQ tests to have any degree of accuracy, every student would have to have the exact same teachers throughout their educational life. And the list goes on. Kaufman's list then makes one think about the analogy: What came first, the chicken or the egg?

Herrnstein and Murray (1994), in their controversial work *The Bell Curve*, spend a great deal of time on the impact of IQs on the lives of individuals. A brief summary of some salient points is apropos here. Let's start with their division of classes of IQ, which is divided into five based on standard sociological categories. The intervals are divided at 5th, 25th, 75th, and 95th percentiles of the distribution. "A great deal of interest goes on within the top 20 percent and bottom 20 percent of the population. … Class III, the normals, comprises half of the population. Classes II and IV each comprise 20 percent, and Class V, like Class I, comprise 5 percent." (pp. 120-21)

The graph on page 121 of their book – Defining the cognitive classes (see below) – shows that those in classes I and II – with IQs ranging from 110 to 150 – are the ones college prep is geared toward and they comprise only 25% of the population yet the vast majority of resources is expended on this population. This suggests that 75% of the population is ill prepared for life due to such prejudice. Even though the bottom 20% receives a great deal of attention, it is not the type of attention that provides for an independent and successful life. They are simply coddled, but without much substantive benefits.

[13] IQ tests such as Cattell's Culture Fair and Raven's Progressive Matrices tests measure some spatial abilities and are independent of cultural experiences. However, spatial abilities are gifts no different than musical or athletic or mathematical or linguistic abilities. It is simply one of several human abilities.



Next, Herrnstein and Murray (p. 146) compare IQs to high school completion:

Failure to Get a High School Education Among Whites

Cognitive Class	Percentage Who Did Not Graduate or Pass a High School Equivalency Exam
I – Very bright	0 [*of very bright population*]
II – Bright	0.4 [*of bright population*]
III – Normal	6 [*of normal population*]
IV – Dull	35 [*of dull population*]
V – Very dull [*IQs 75 & under*]	55 [*of very dull population*]
Overall average	9

The "very dull" and "dull" are the populations at the greatest risk and have the most social problems, and yet we do nothing for them but provide the penitentiary system to keep them restrained once they are through with school and find it impossible to fit in. Or if they don't end up in jail, they may: frequent our hospitals with health problems due to poor care; frequent our judicial halls for various infractions; or

collect unemployment, child support, or welfare checks and then vote for big government politicians to keep the welfare spigot open. So much of this could be avoided with an appropriate education system.

There are so many variables that determine IQ – such as the optimization of the system for particular talents and cultural experiences – that statistics like these masks a multitude of them and therefore little can be gleaned from such figures. One thing that becomes obvious, however, is that a large percentage of individuals are marginalized since those with IQs lower than 110 typically get little return on the public's investment. Therefore, Herrnstein and Murray's figures show us that other types of tests that suit other intelligences are required so other professions can be made available to the various school age populations. Aptitude tests capture other intelligences and therefore should come to dominate the educational assessment field. IQ tests might be better used in those fields that require outstanding memory and recall. Think of pharmacology as an example of such requirements.

Herrnstein and Murray speak very highly of the Armed Forces Qualification Test (AFQT). They state, "The AFQT is extracted from the scores on several tests that everyone in the armed forces takes. … The resulting database has no equal in the study of job productivity. … [T]he results from the military conform to the results in the civilian job market." (p. 73)

I remember taking this battery of tests and have to say I was very impressed with them. Applied Education Foundation researchers will have to delve deep into this to see what civilian public education systems can gain from the military's practical approach to reality.

Related to the Military's AFQT assessment battery is Professor John Holland's vocational testing named the Holland Codes. Professor Richard Haier (in Peterson interview, 2017) points out that Holland's work at John Hopkins University was "very powerful."

The graph on page 149 of *The Bell Curve* is titled "In predicting which white youths will never complete a high school education, IQ is more important than socioeconomic status (SES)"; then it states "Probability of permanently dropping out of high school" which shows "As IQ goes from low to high" the probability of quitting school goes from approximately 62% to near 0; and then for SES it goes from 20% to 0. This demonstrates that the public school system offers little to nothing for those with low IQs. Once again, what comes first the chicken or the egg? Does a low IQ reflect mental incapacity or does it show the results of a poor education or one unfitting of the individual? I would say, on average, it is unfitting of individual needs.

Herrnstein and Murray share what critics had to say about the dominance the IQ tests were playing in the lives of citizens and the effects on society:

> Young Walter Lippmann, already an influential columnist, was one of the most prominent [critics], fearing power-hungry intelligence testers who yearned to "occupy a position of power which no intellectual has held since the collapse of theocracy." In a lengthy exchange in the *New Republic* in 1922 and 1923 with Lewis Terman, premier American tester of the time and the developer of the Stanford-Binet IQ test, Lippmann wrote, "I hate the impudence of a claim that in fifty minutes you can judge and classify a human being's predestined fitness in life. I hate the pretentiousness of that claim. I hate the abuse of scientific method which it involves. I hate the sense of superiority which it creates, and the sense of inferiority which it imposes." … As

> Lippmann feared, people did tend to give more credence to an individual's specific IQ score and make broader generalizations from it than was appropriate. (p. 6)

> Walter Lippmann's hostility toward intelligence testing was grounded in his belief that this most important of all human qualities was too diverse, too complex, too changeable, too dependent on cultural context, and, above all, too subjective to be measured by answers to a mere list of test questions. Intelligence seemed to him, as it does to many other thoughtful people … more like beauty or justice than height or weight. Before something can be measured, it must be defined…. (p. 17)

There are those who seriously question the accuracy and effectiveness of assessment tests and whether they highly favor certain social biases.

> Today, voices for the elimination of standardized tests are few. One is Linda S. Siegel, professor in the Department of Educational Psychology and Special Education at the University of British Columbia in Vancouver, Canada. … According to most definitions – although they are not conclusive – intelligence is made up of the skills of logical reasoning, problem solving, critical thinking, and adaptation. This scenario seems reasonable, until one examines the content of IQ tests. The definition of intelligence, as is operationalized in all IQ tests, includes virtually no skills that can be identified in terms of the definitions of intelligence. To support her statement, Siegel gives a detailed analysis of the subtests of the *Wechsler Intelligence Scale for Children-Revised* (WISC-R). This IQ test is composed of Verbal and Performance sections, and is nearly always used in LD diagnosis. In each subtest of the Verbal scale, performance is in varying degrees dependent on specific knowledge, vocabulary, expressive language and memory skills, while in the Performance scale, visual-spatial abilities, fine motor coordination, perceptual skills, and in some subtests speed, are essential for scoring. As Siegel rightly points out, IQ tests measure, for the most part, what a person has *learned*, not what he is capable of doing in the future (his potential). (Edublox)

Siegel points out that as it relates to the LD student, the IQ test may measure a weakness in one area, but then due to the low score in this one area, the student's IQ score places him in an overall LD category, thereby marginalizing him.

> The unreliability of IQ tests has been proved by numerous researchers. The scores may vary by as much as 15 points from one test to another, while emotional tension, anxiety, and unfamiliarity with the testing process can greatly affect test performance. In addition, Gould described the biasing effect that tester attitudes, qualifications, and instructions can have on testing. In one study, for example, ninety-nine school psychologists independently scored an IQ test from identical records, and came up with IQs ranging from 63 to 117 for the same person. (Edublox)

Gardner (1983) references the Swiss psychologist Jean Piaget who was

> interested in the errors children make when tackling items on an intelligence test. Piaget came to believe that it is not the accuracy of the child's response that is important, but rather the lines of reasoning the child invokes: these can be most clearly seen by focusing on the assumptions and

> the chains of reasoning that spawn erroneous conclusions. … [I]n looking at the scientific moves [Piaget] made, one can gain a feeling for some of the inadequacies of the Binet-Simon program. First of all, the I.Q. movement is blindly empirical. It is based simply on tests with some predictive power about success in school and, only marginally, on a theory of how the mind works. There is no view of process, of how one goes about solving a problem: there is simply the issue of whether one arrives at a correct answer. For another thing, the tasks featured in the I.Q. test are decidedly microscopic, are often unrelated to one another, and seemingly represent a "shotgun" approach to the assessment of human intellect. The tasks are remote, in many cases, from everyday life. …
>
> Much of the information probed for in intelligence tests reflects knowledge gained from living in a specific social and educational milieu. … In contrast, intelligence tests rarely assess skill in assimilating new information or in solving new problems. This bias toward "crystallized" rather than "fluid" knowledge can have astounding consequences. … To put it in the terms of the Soviet psychologist Lev Vygotsky, intelligence tests fail to yield any indication of an individual's "zone of potential or 'proximal' development." (pp. 17-18)

Jensen (1969, p. 13) explains fluid and crystallized "intelligence":

> Raymond B. Cattell (1963) has made a conceptually valid distinction between two aspects of intelligence, *fluid* and *crystallized*. … *Fluid* intelligence is the capacity for new conceptual learning and problem solving, a general "brightness" and adaptability, relatively independent of education and experience, which can be invested in the particular opportunities for learning encountered by the individual in accord with his motivations and interests. … *Crystallized* intelligence, in contrast, is a precipitate out of experience, consisting of acquired knowledge and developed intellectual skills. Fluid and crystallized intelligence are naturally correlated in a population sharing a common culture, because the acquisition of knowledge and skills in the first place depends upon fluid intelligence. While fluid intelligence attains its maximum level in the late teens and may even begin to decline gradually shortly thereafter, crystallized intelligence continues to increase gradually with the individual's learning and experience all the way up to old age.

Peterson (2017, 53 minutes into the interview) Dr. Richard Haier discussed brain imaging in an attempt to better understand "intelligence." By imaging volunteers' brains during active learning, as well as after mastery had taken place, Haier found that the primary areas of the brain that were associated with cognitive processes were related to *language*, *memory*, and *attention*, which he felt provide the "architecture on which intelligence is built." One may surmise that highly refined *linguistic* abilities provide the means to understand and express ideas efficiently and effectively, which "intelligence" tests are highly loaded for. Tests are also highly loaded for *memory*, especially rapid recall abilities. And *attention* is necessary to accomplish just about anything, be it passive learning or active application of something learned.

While *memory* abilities are in my humble opinion fixed, *linguistic* abilities can be improved, even though such abilities differ between people. I think the same can be said for *attention*. Though it differs between people, it can be improved. When I taught martial arts to children, parents noticed marked improvements in their children.

* * *

A young Australian lady, who creates science videos and goes by the name Tibees, created a video titled *This is What a Synesthesia Test Looks Like*. Let's look at this human attribute to see how it may affect assessment testing.

Since memory and recall are improved by associations, perhaps those with synesthesia talents have an advantage on IQ tests. "Richard Cytowic, a well-known neurologist and author, brings us what is likely the best info-per-second video on synesthesia in his TED-Ed, "What Color is Tuesday?" … Cytowic points out that there is nothing wrong with the synesthetic mind; in fact, it is likely to possess a better capacity for memory and recall." He states that synesthesia endows synesthetes with superior memory and that synesthesia is in 4% of the population. (Also, *see* Synesthesia Test.)

This points to one of a multitude of differences between humans. Each human talent/attribute provides an advantage in certain realms over those who do not adequately possess such an attribute, in relative terms. Perhaps Howard Gardner should add synesthesia to his list of multiple intelligences; and perhaps we should, once and for all, unseat IQ's lofty position and reduce it to the same level of all other human attributes. This would go a long way in establishing a more equitable society through natural policies rather than through authoritarian regulatory mechanisms.

* * *

To conclude this section, I would like to offer the following which was used in my essay *Transfer of Learning*, but is just as important in this essay on academic assessment:

Mestre et al. (2005) compares reasoning abilities to memory abilities showing how they are decoupled:

> **[O]ne consistent finding is that reasoning accuracy is independent of memory accuracy across a broad range of problems: Memory for problem facts was found to be unrelated to reasoning about those same facts. This finding contradicts theories that rely on assumptions of limited working memory as an explanatory construct in reasoning and problem solving. (p. 54) … Under certain circumstances, factors that enhance verbatim memory [memory for surface details] actually depress reasoning performance**. (p. 55) (Emphasis added)

This flies in the face of everything academic today! The educational community believes that memory and reasoning go hand in hand and assessment tests, which includes IQ tests and which are correlated with high stakes outcomes, reflect this belief. Indeed it is a belief, not a truth, with devastating consequences to those who may have outstanding reasoning abilities, but who lack outstanding memory abilities. The harm done to individuals and society is immeasurable!

If foundational concepts were offered to students that provide overarching ideas, rather than disconnected detailed data that are memorized for tests, such as rules[14], everyone would more readily experience transfer. Literacy and numeracy would improve dramatically. There may be those in the academy who would bemoan the loss of the ability to rank students since outcomes across populations would actually experience a leveling effect. However, in the place of a ranking system that seeks to marginalize a majority of citizens, assessment methodologies could be used to discover talents in order to steer each individual in a direction that affords social success in its various manifestations.

Algorithmic Thinking versus Rational Thought

To begin this chapter, let's consider what a human biologist, Gary Brecka (in an interview with Steven Bartlett[15]), has to say about his photographic memory:

> Sometime in the 8th grade, I discovered I was clinically photographic. So, I have a clinical level of photographic recall, which is different than having a visual memory, so I have a voluminous capacity to recall things that I read – even if I don't understand them, which is why I never read for pleasure, because I record everything. So, I can fill my brain with senseless things and record it and regurgitate it, or I can fill it full of things that fascinate me, so I naturally gravitated toward science. … When I read peer reviewed papers and scientific journals, it doesn't make me more intelligent than someone else because very often I can recall information that I don't understand. So, I can regurgitate voluminous amounts of information. Subjects that rely on rote memorization like chemistry, biology, neurobiology, microbiology, a lot of these sciences don't actually make sense – you just have to memorize how they operate. Chemistry in a lot of ways, doesn't make sense. You just have to remember what happens when you put "these" two elements together, they create "this."

At the end of the 19th century, the German Statist movement was embraced by a sector of American society which saw the academic community as the means to further their interests. They donned the moniker *Progressivism* to denote the concept of progress toward a Statist utopian society. They used social engineering methodologies to achieve these goals. Education was at the foundation of their plans since indoctrination is easiest to achieve during the formative years of individual's. Therefore, the entire educational system required a complete overhaul.

Amongst various strategies necessary to achieve radical change, *Social Efficiency* was one tactic to reach the utopian dream. This meant that teaching and assessment strategies were designed to maximize efficiencies of the desired goal, based on the principle of the greatest good for the greatest amount of people. This required sacrificing traditional educational methodologies, which had embraced the development of rational thought in conjunction with the development of high levels of literacy and numeracy understanding to provide the opportunity for every individual to

---

[14] Let's consider grammatical *rules* as an example. If one were to read the works of great minds over the centuries, it would quickly become apparent that grammar was all over the board. Yet, nothing is lost in understanding the intent of these great minds. However, today, if one doesn't follow grammatical *rules*, doors may very well be shut to future educational opportunities even though an individual's mind might be quite profound. Therefore, instead of *rules* being dictated that must be memorized, it would be better to teach *principles* of language found in the science of linguistics. If one becomes a journalist, for example, let the publisher provide the *rules* it operates under. With a firm understanding of *principles*, an understanding of *rules* is easier to acquire if and when it is needed.

[15] E225: The Man Who Predicts How Long You Have to Live (To The Nearest Month): Gary Breka.

achieve his greatest good. The contrast between traditional American culture and this new Statist religion could not be more glaring.

Here we observe the split between the two political camps that now dominate the West: The individualists who saw social advancement being achieved through the improvement of individual virtue, which originated in Stoicism and Christianity; and the collectivists who desired social perfection through manipulation and control of the masses (i.e. Statism, Progressivism, Fascism, Socialism, and Marxism). At the end of the 19th century, collectivism came to dominate which led to the bureaucratization of the educational system and the establishment of the one-size-fits-all teaching and testing methodologies.

New "scientific" instructional methodologies were adopted from Germany to inculcate Statist ideals of obedience to the State first and foremost with the secondary purpose of developing good compliant workers with sufficient fundamental language and computational abilities to work effectively in institutional environments.

Progressivism embraced the concept of sorting people according to their abilities for potential roles in the Statist social hierarchy. To determine whether the new teaching methods were effectively achieving their goals, assessment methodologies became a necessity – enter IQ testing and a shift from the development of rational thought to the development of algorithmic thinking. In spite of numerous insightful psychologists' objections to this ill-conceived and dangerous "intelligence" testing regime, it has been deified in the education culture since the turn of the last century. However, in the last few decades, numerous cognitive psychologists have been chipping away at IQ's sacred status and slowly but surely, they are tearing down the fortress.

Stanovich (2009) is an example of a cognitive psychologist who objects to the deification of the perceived value of IQ sorting and IQ tests, and their proxies such as the SAT and ACT. He offers a perspective that may very well lift the veil of murkiness that hides the Progressives' schemes and methods designed to radically alter the American political landscape in order to achieve the "perfect society." In order to show the reader how Stanovich incidentally opens this window to view this landscape, we must analyze what he has to say about what IQ tests offer and what they miss.

While Stanovich argues IQ tests do offer insight into certain cognitive abilities, they only measure algorithmic thinking abilities; they do not measure rational thought. Consequently, the education establishment does not teach people to think rationally, but rather, it teaches students to follow instructions well, in conjunction with memory and recall abilities, and then it selects people – who demonstrate these abilities – for further education and lucrative credentialing for the desired Statist hierarchy.

Stanovich exposes the contrast between algorithmic thinking and rational thought by revealing a common dilemma that confuses most people: How is it "smart" people (people with high IQs) do stupid things?

> [T]he way that we have historically measured intelligence makes this phenomenon not perplexing at all. If by smart we mean IQ-test smart and by dumb we mean poor decision making, then the source of the phenomenon is clear. IQ tests do not measure adaptive decision

> making. So, if we are surprised at a high-IQ person acting foolishly, it can only mean that we think that all good mental attributes must co-occur with high intelligence – in this case, that rational thinking must go with high intelligence. However, research is increasingly bringing this assumption into question.

Stanovich points out that what IQ tests measure is a narrow field of abilities that he associates with algorithmic cognitive structures. If this is true, and I believe it is, then what IQ tests assess is fundamentally the same abilities that computers are designed for. There is a direct and powerful correlation between what our education establishment pushes for and the abilities of computers. Perhaps this is why some believe that computers can be designed for artificial intelligence that will rival human brainpower.

In further distinguishing algorithmic processing from rational thought, Stanovich states:

> Importantly, the algorithmic mind can be evaluated in terms of efficiency but not rationality. This concern for the efficiency of information processing as opposed to its rationality is mirrored in the status of intelligence tests. They are measures of efficiency but not rationality…. All tests of intelligence or cognitive aptitude are optimal performance assessments, whereas measures of critical or rational thinking are often assessed under typical performance conditions.
>
> The difference between the algorithmic mind and the reflective mind is captured in another well-established distinction in the measurement of individual differences – the distinction between cognitive abilities and thinking dispositions. The former are … measures of the efficiency of the algorithmic mind. The latter travel under a variety of names in psychology – thinking dispositions or cognitive styles being the two most popular. Many thinking dispositions concern beliefs, belief structure, and, importantly, attitudes toward forming and changing beliefs. Other thinking dispositions that have been identified concern a person's goals and goal hierarchy. Examples of some thinking dispositions that have been investigated by psychologists are: actively open-minded thinking, need for cognition (the tendency to think a lot), consideration of future consequences, need for closure, superstitious thinking, and dogmatism.
>
> … It is only necessary to note that the types of cognitive propensities that these thinking disposition measures reflect are: the tendency to collect information before making up one's mind, the tendency to seek various points of view before coming to a conclusion, the disposition to think extensively about a problem before responding, the tendency to calibrate the degree of strength of one's opinion to the degree of evidence available, the tendency to think about future consequences before taking action, the tendency to explicitly weigh pluses and minuses of situations before making a decision, and the tendency to seek nuance and avoid absolutism. In short, individual differences in thinking dispositions are assessing variation in people's goal management, epistemic values, and epistemic self-regulation – differences in the operation of the reflective mind. They are all psychological characteristics that underpin rational thought and action.

The manner in which the author lays out the differences between the rational mind and algorithmic mind reveal the differences between leaders and followers. Highly effective leaders must have all the attributes of the reflective mind as stated by the author in his above list of "thinking dispositions."

Whereas, followers are required to be efficient in getting a job done that has been assigned to them by a leader, which requires algorithmic abilities and which is what our educational system is exclusively designed to help develop and then measure. This reveals the need to differentiate between such abilities and consequently to educate individuals based on their own unique abilities.

Let's use a hypothetical business arrangement of creating and assembling puzzles to demonstrate an analogy that helps clarify my point. The rational thinking person would be the one to paint a picture, and then design the shapes of the puzzle pieces for workers to then produce. Once the puzzle has been created, those who possess algorithmic talents will be the ones who can produce it the quickest, and therefore will be the ones who will be hired. While this may be a simplistic analogy, it reveals, at the fundamental level, what our education system promotes – good worker bees. Granted, there are different levels of responsibilities for workers, so those with high IQs will be assigned the more complicated tasks; but they are tasks, not creative, diagnostic, and innovative efforts.

Dysrationalia:
The term "dysrationalia" was coined by Stanovich. It can be used to draw attention to what is missing in IQ tests, what's missing from our educational system, and consequently, what is missing from the cognitive abilities of so many Americans given the restraints they endured during the formative years.

> I define dysrationalia as the inability to think and behave rationally despite having adequate intelligence.
>
> … One of the reasons that many of us are dysrationalic to some extent is that, for a variety of reasons, we have come to overvalue the kinds of thinking skills that IQ tests measure and undervalue other critically important cognitive skills, such as the ability to think rationally.
>
> … IQ tests are good measures of how well a person can hold beliefs in short-term memory and manipulate those beliefs, but they do not assess at all whether a person has the tendency to *form* beliefs rationally when presented with evidence. And again, similarly, IQ tests are good measures of how efficiently a person processes information that has been provided, but they do not at all assess whether the person is a *critical assessor* of information as it is gathered in the natural environment.

I think this is an outstanding analysis of how our education system is designed, which IQ tests reflect: It promotes short-term memory abilities (necessary to be a good worker-bee) of prepackaged information handed to an individual, but ignores the abilities to digest and independently formulate conclusions that are grounded in sound reasoning (necessary for self-determined individuals). Stanovich then makes the point that "IQ tests determine, to an important degree, the academic and professional careers of millions of people in the United States." I would call this situation an unmitigated disaster.

Stanovich further explains what's missing from IQ tests:

> The cognitive abilities assessed on intelligence tests … are not about high-level personal goals and their regulation, or about the tendency to change beliefs in the face of contrary evidence, or about how knowledge acquisition is internally regulated when not externally directed. As we shall see …, people have indeed come up with *definitions* of intelligence that encompass such

things. Theorists often define intelligence in ways that encompass rational action and belief but, despite what these theorists argue, *the actual measures of intelligence in use assess only algorithmic-level cognitive capacity*. No current intelligence test that is even moderately used in practice assesses rational thought or behavior.

… [A]ssessments of [rational] thinking are nowhere to be found on IQ tests. Adaptive decision making is the quintessence of rationality, but the items used to assess intelligence on widely accepted tests bear no resemblance to measures of rational decision making.

Judgment and decision-making skills – the skills of rational thought – are at least as important as the attributes that are assessed on IQ tests. Like intelligence, rational thinking skills relate to goal achievement in the real world. Yet we fail to teach them in schools or to focus our attention on them as a society. Instead, we keep using intelligence proxies as selection devices in educational institutions from exclusive preschools to graduate schools. Corporations and the military are likewise excessively focused on IQ measures. The lavish attention devoted to intelligence (raising it, praising it, worrying when it is low, etc.) seems wasteful when we virtually ignore another set of mental skills with just as much social consequence.

Intelligence tests are … radically incomplete as measures of cognitive functioning. Because of their vast influence, IQ tests have both explicitly and implicitly defined, for the layperson and psychologist alike, what cognitive attributes to value. These are important abilities, to be sure, but the tests leave out huge domains of cognitive functioning. … The skills of judgment and decision making are cognitive skills that are the foundation of rational thought and action, and they are missing from IQ tests. … [S]tandard IQ tests devote no section to rational thinking as cognitive scientists would define the term.

Broad Theories and Narrow Theories of Intelligence:
Stanovich covers the divide that has become entrenched between those loyal to the use of the word "intelligence" as psychometricians understand it, and those who use it to denote powerful mental cognitive capacities.

[W]e are bumping up against an old controversy in the study of cognitive ability – the distinction between broad and narrow theories of intelligence. Broad theories include aspects of functioning that are captured by the *vernacular* term *intelligence* (adaptation to the environment, showing wisdom and creativity, etc.)…. Narrow theories, in contrast, confine the concept of intelligence to the set of mental abilities actually tested on extant IQ tests.

… The so-called folk language (everyday usage) of the term *intelligence* is an utterly inconsistent mess. It is a unique confluence of inconsistent terminology, politically infused usage, and failure to assimilate what science has found out about the nature of human cognitive abilities.

… It might help the discussion of broad versus narrow views [of "intelligence"] if we mark these abilities with an easily remembered acronym – MAMBIT (to stand for: the mental abilities measured by intelligence tests). The narrow view of the intelligence concept, in viewing intelligence as MAMBIT, differs from the broad view in expressly *not* including in its primary

definition a host of things that appear in broad theories: adaptation to the environment, real-life decision making, showing wisdom and creativity, etc.

… In short, the folk theory overvalues MAMBIT by viewing as odd any case where other good mental qualities do not go together with high MAMBIT. In this way, the folk theory undervalues other mental faculties by giving pride of place to MAMBIT….

… Broad theories conjoin the two (MAMBIT and sensible decision making) under the umbrella term *intelligence*. Such broad views of intelligence lead to the privileging of MAMBIT and the devaluing of the non-MAMBIT parts of the broad definition. This is because MAMBIT has a name (IQ), is measured explicitly (by IQ tests), and has a one-hundred-year history that many people know at least a little about. If we would name (and measure) the other things (and not just call them part of intelligence), we would be better able to give them proper emphasis. And we do have an omnibus name for these other things. Adaptive behavioral acts, judicious decision making, efficient behavioral regulation, sensible goal prioritization, reflectivity, the proper calibration of evidence – all of the characteristics that are lacking when we call an action foolish, dumb, or stupid – are precisely the characteristics that cognitive scientists study when they study rational thought.

Most psychologists realize that IQ tests do not encompass all of the important mental faculties. Most educators also would know this if asked explicitly. Yet despite this, I still contend that *most of the time, most people forget this fact*. In short, I think that IQ tests do fool most of the people most of the time – including psychologists who should know better. By acknowledging the frequent occurrence of dysrationalia, we create the conceptual space to value abilities at least as important as MAMBIT – abilities to form rational beliefs and to take rational action.

… Professional psychologists will immediately recognize my proposal to identify intelligence only as MAMBIT as a version of E. G. Boring's infamous dictum – and this recognition may cause some of them to balk at my proposal. Boring's dictum was that we should define intelligence as what the intelligence tests measure. However, what made Boring's suggestion objectionable was that neither he nor anyone else at the time (1923) knew what the tests measured. Because of this, Boring's definition of intelligence was truly circular. The situation now is totally different. We now know – from the standpoint of information processing and cognitive neuroscience – what the tests measure.

… Perhaps some of this attention to intelligence is necessary, but what is not warranted is the ignoring of capacities that are of at least equal importance – the capacities that sustain rational thought and action. It is ludicrous for society to be so fixated on assessing intelligence and to virtually ignore rationality when it is easy to show that the societal consequences of irrational thinking are profound. And yet, oddly enough, I have discovered that there is enormous resistance to the idea of giving full value to mental abilities other than intelligence.

Of course, currently, we do not have a rationality quotient, as we have an intelligence quotient, which might explain, at least to some extent, why IQ has acquired such value in relation to other equally important cognitive skills. In our society, what gets measured gets valued.

> … If not for professional inertia and psychologists' investment in the IQ concept, we could choose tomorrow to more formally assess rational thinking skills, focus more on teaching them, and redesign our environment so that irrational thinking is not so costly.
>
> … Finally, my argument is, essentially, that we would value MAMBIT less if we would take care to label the things it is not (rationality) and not let the term *intelligence* incorporate those other things.

The author wishes to utilize the word "intelligence" in an exclusive association with MAMBIT, but I believe he is swimming against the current. The word "intelligence" has been genericized to mean all human cognitive abilities associated with good mental capacities and abilities. To make an analogy with consumer product marketing, the word "intelligence" does not possess the trademark Stanovich is attempting to brand it as. Therefore, it is not possible to expect the public at-large to adopt his narrow definition of the word. IQ should be identified as algorithmic thinking while all other cognitive abilities need their own identities; not so unlike what Howard Gardner has provided with his theory of *multiple intelligences*. The sum of all cognitive abilities is *wisdom*; but wisdom is rooted in the pursuit of virtue – a never ending endeavor.

To bring perspective into the analysis of what IQ lacks, Stanovich points out that other mental abilities actually provide superior results over high IQs

> In an important study, [it was] found that the grade point averages of a group of eighth graders were predicted by measures of self-discipline…. A longitudinal analysis showed that self-discipline was a better predictor of the changes in grade point average across the school year than was intelligence. The personality variable of conscientiousness – which taps the higher-level regulatory properties of the reflective mind – has been shown to predict, independent of intelligence, academic performance and measures of performance in the workplace.
>
> … [T]he reflective mind not only accesses general knowledge structures but, importantly, accesses the person's opinions, beliefs, and reflectively acquired goal structure. The algorithmic mind accesses micro-strategies for cognitive operations and production system rules for sequencing behaviors and thoughts.

Crystalline versus fluid cognition[16]:
Stanovich states,

> … It is true that much of the mindware of rational thought would be classified as crystallized intelligence in the abstract. But is it the kind of crystallized knowledge that is specifically assessed on the tests? The answer is no. The mindware of rational thought is somewhat specialized mindware (it clusters in the domains of probabilistic reasoning, causal reasoning, and scientific reasoning…). In contrast, the crystallized knowledge assessed on IQ tests is

---

[16] Fluid intelligence is the ability to use logic and solve problems in new or novel situations without reference to pre-existing knowledge. Crystallized intelligence is the ability to use knowledge that was previously acquired through education and experience. https://www.thoughtco.com/fluid-crystallized-intelligence-4172807

deliberately designed to be nonspecialized. The designers of the tests, in order to make sure the sampling of Gc [crystalline *g*] is fair and unbiased, explicitly attempt to *broadly* sample vocabulary, verbal comprehension domains, and general knowledge. The broad sampling ensures unbiasedness in the test, but it inevitably means that the specific knowledge bases critical to rationality will go unassessed. In short, Gc, as traditionally measured, does not assess individual differences in rationality, and Gf [fluid *g*] will do so only indirectly and to a mild extent.

Processing versus Content Limitations: Stanovich provides,

> [T]he human brain is characterized by two broad traits that make it less than rational – one a processing problem and one a content problem. The processing problem is that we are cognitive misers. The content problem comes about because we need to acquire some very specific knowledge structures in order to think and act rationally. When knowledge structures that are needed to sustain rational behavior are not present, I will term this a mindware problem….
>
> … The development of probability theory, concepts of empiricism, logic, and scientific thinking throughout the centuries have provided humans with conceptual tools to aid in the formation and revision of belief and in their reasoning about action. They represent the cultural achievements that foster greater human rationality when they are installed as mindware. As societies evolve, they produce more of the cultural tools of rationality and these tools become more widespread in the population.
>
> … The tools of rationality – probabilistic thinking, logic, scientific reasoning – represent mindware that is often incompletely learned or not acquired at all. This incomplete learning represents a class of causes of irrationality that I label a "mindware gap." … In fact, some acquired mindware can be the direct cause of irrational actions that thwart our goals. This type of problem I term "contaminated mindware."
>
> … It is often difficult to specify what the best type of performance might be, but performance errors are much easier to spot. Essayist Neil Postman has argued, for instance, that educators and other advocates of good thinking might adopt a stance more similar to that of physicians or attorneys. He points out that doctors would find it hard to define "perfect health" but, despite this, they are quite good at spotting disease. Likewise, lawyers are much better at spotting injustice and lack of citizenship than defining "perfect justice" or ideal citizenship. Postman argues that, like physicians and attorneys, educators might best focus on instances of poor thinking, which are much easier to identify, as opposed to trying to define ideal thinking.[17]
>
> … To jointly achieve epistemic and instrumental rationality, a person must display judicious decision making, adequate behavioral regulation, wise goal prioritization, sufficient thoughtfulness, and proper evidence calibration. … People fail to fulfill the many different strictures of rational thought because they are cognitive misers, because they lack critical mindware, and because they have acquired contaminated mindware. These errors can be

[17] The fact that Postman developed this concept, in all likelihood, is related to Progressives' attempt to reengineer society to achieve utopian perfection. They have been attempting to inculcate the "perfect" Statist way of thinking since they appropriated the education system.

> prevented by acquiring the mindware of rational thought and the thinking dispositions that prevent the overuse of the strategies of the cognitive miser.

This can best be achieved by returning to the pre-Progressive ideas of educational foundations being grounded in classical studies based on the development of reasoning abilities.

Type 1 versus Type 2 Mental Processing:
Stanovich analyzes the human tendency (which is found in all animals) to be as efficient as possible in the use of brain activity. He speaks of Type 1 (quick and superficial) and Type 2 (deep and analytical) mental processing and how, by default, our brains resort to Type 1 as often as possible, and which frequently leads to less than rational outcomes. He then suggests,

> But modern life often requires more precise thought than this. Modern technological societies are in fact hostile environments for people reliant on only the most easily computed automatic response. … But why are we cognitive misers and as a result less than fully rational? In a word – evolution. Our cognitive mechanisms were designed by evolution, and evolution does not operate to produce humans who are perfectly rational.

This is why education to achieve this end is sooooo important! Again, this was the goal of the classical education program.

Deification of Intelligence:

Stanovich argues against the veneration of high IQs, or more appropriately labeled, *algorithmic processing abilities*.

> … The Dark Side of the Deification of Intelligence: Such deification of intelligence can have a truly perverse moral consequence that we often fail to recognize – the denigration of those low in MAMBIT. Such denigration goes back to the very beginnings of psychometrics as an enterprise.
>
> … The historical tendency that Robert Sternberg has noted – the "tendency to conflate scores on tests of intelligence with some kind of personal value" – appears in modern life in many guises. As Sternberg suggests, intelligence has come to signify something like one's personal essence – some indication of personal worth. The deification of IQ and the denigration of low intelligence is now so complete that people would rather have a high IQ than almost any other physical or mental quality.

This reveals a lack of true intelligence of those who designed and have bought into the whole IQ culture. Since this deification of IQ does great harm to most in society, how can its dominance be seen as a social good, and since it is not a social good, it reveals a stupidity (dysrationalia) of those who embrace it.

Stanovich unintentionally provides an analysis of the transition from a classical education system to the Progressive/Statist model that sought to perfect society through science and the bureaucratic power of government. Much of what he points to in comparing the algorithmic versus the rational mind processes reveals what happened when Progressives took over education. Previous to their arrival, education was

geared to developing the rational and reflective mind. The classic works that were studied, were geared toward this primary goal. But with the Industrial Revolution shifting societal economic forces away from agriculture and cottage industries and toward institutional business as well as bureaucratic arrangements, the Progressive movement took over education and shifted it from the development of rational thinking, as demonstrated in the classical works that previously dominated a good education, and more toward algorithmic thinking that is dependent on facts, data, memory, recall, and speed of work, that creates good and obedient workers who follow orders well and do not question things. A good example of this is the fact that many teachers and most professors discourage questions from students. They simply want students to listen and follow. Stanovich alludes to this when he provides: "In short, our research converges with that of other researchers in indicating that in informal reasoning situations where people are not told to put aside their prior beliefs, intelligence is unrelated to the tendency to reason in an unbiased manner." This demonstrates that the education system is indeed designed for indoctrinating a bias and sorting out obedient workers for institutional environments.

With Progressivism's methods of ruling society, the rational mind was seen as a threat since it bred independent thinking (which Progressives identified in a condescending way as "rugged individualism") which is not tolerable under Statist social engineering strategies. Therefore, education became the instrument to mold compliant and obedient workers to achieve the "greater good" of society, with the expectation of its eventual "perfection" or utopia as they put it.

When classic studies were the order of the day and IQ testing methods had not been invented yet, essay writing was used extensively to assess rational thought. Generally speaking, showing one's methods of coming to a conclusion on something was the means to determine a student's abilities and understanding of cognitive processes. In other words, such assessment methodology doesn't necessarily require right-wrong type of analysis since we can never be completely cognizant of all variables on a given subject. What's most important is *analysis* itself. Whatever conclusions an individual may come to, what matters most is the method of arriving at a conclusion. In other words, it's the journey rather than the destination; or the means rather than the ends. This is so because most conclusions are not simple black and white truths. There are subtle nuances that effect truths which requires a blend of logic and intuition if wisdom is to prevail. And make no mistake, wisdom is far more important than what passes for "intelligence" in contemporary academe.

We need to incorporate classical studies in education once again since they are deeply rooted in the idea of building on the very structures the author champions. See Appendix to observe a traditional classical education offered at St. John's College. It must be born in mind that I am not advocating that everyone attends such an extensive program. Individual needs and abilities will determine how much will be required and at what stage of development it should be introduced. However, I believe the earlier the better. Exposure to such an education should not wait until the postsecondary years when stages of development are coming to a close. The secondary years – 6th through 12th grades – would be best.[18] This was the original timeframe before the Progressives took over and I believe Stanovich's observations demonstrate a deviation from the purpose of education, which is to develop rational thinking abilities rather than good worker bees.

[18] When I home schooled my sons during the high school years, I used classical education as the primary curricula. Now that they are adults and are married with children, it is quite obvious, when compared to other people, that their classical education has served them **very** well. They are far more capable of using intuition and reason when faced with resolving challenges than the vast majority of people they know.

Stanovich's analyses reveal the need to cultivate an education system grounded in the pursuit of developing reason in individuals, which, again, was the goal of the classical education program. We have abandoned these studies at our peril. Socialist utopian dreams now cloud the minds of illiterates who actually pass for highly educated people. And it seems the more time spent in educational circles, the greater is their delusional perspective, as exemplified by their adoption of Fascist, Socialist, or Communist revolutionary efforts (perhaps the archetypical example of dysrationalia). Instead of providing reasoned arguments to convince others of their utopian ideas, collectivist types use deception to destroy Western culture so they may rebuild society in the likeness of **their** "ideal." Education, which is the means to this end, now indoctrinates in the ways of chaos and madness through the pursuit of omnipotent power, which is a recipe for disaster as it leads to absolute corruption.

High Schools and High-Stakes Testing

Carnoy et al. (2003) analyze the standards-based accountability reform policies for high schools that had their beginning in the late 20th century. Their book "makes the important conclusion that, without corresponding internal structures of coherence and accountability, externally imposed educational reforms make little or no difference in the quality of education, or in improving high school graduation rates." (Back cover) The appropriate organizational culture is crucial for success.

The most important lesson discovered by the researchers in this work is expressed by Richard Elmore:

> [T]he capacity of the school to deliver high quality education is still the fundamental measure of how a school responds to the increased pressures exerted by the new state accountability systems. Elmore argues that testing schools and rewarding and sanctioning them can only take schools so far in self-improvement. Without the internal coherence (the internal capacity) to make such improvements, schools are less likely to make the organizational changes required for their students to do better as a result of higher standards imposed by states. (p. 11)

The first chapter of the book starts off with the authors explaining the changes that occurred in the late 20th century:

> Today's assessment-based reforms are hardly new, but the purpose of assessment has been "transformed" from traditional goals of measuring intelligence, tracking students, standardizing learning, and evaluating applicants into new forms of judging the quality and equality of schooling. Assessment has also been greatly intensified. The convergence of these two trends has led to more state testing and new types of assessment, such as high stakes tests for high school graduation, and the use of assessment in formal accountability systems. (p. 13)

Chapter 3 begins by explaining the evolution of education throughout the first half of the 20th century. The challenge educators faced were enrollment growth and the heterogeneity of students. At the turn of the last century, higher levels of education was the province of those with academic talents, but with the drastic increases in enrollment of larger percentages of the population, other talents had to be taken into account which required a broader educational offering that some educators have disparagingly referred to as "a shopping mall" of product offerings. Academic talent was, and still is, perceived as "superior,"

so offerings for this sector tended to maintain "high standards." However, for those without this talent, another standard was required.

> Just what those standards would be, however, were not specified by reformers nor standardized by policy makers. Instead, they were quietly and locally negotiated – classroom-by-classroom, track-by-track, subject-by-subject.
>
> …It is against that pattern of practice that current reformers press, arguing for high standards that will be the same for all schools, and for all students within a school. (pp. 87-88)

Here we see the coalescence of the one-size-fits-all approach with the "college for all" mindset. The "shopping mall" product offerings are not acceptable to some reformers since they are perceived as offering inferior products and therefore deny access to become part of the "elite ruling class." Absolutely everyone from lower socioeconomic sectors must be given the opportunity to become part of the ruling class no matter what the cost. The ruin of the majority of students' future, due to the denial of other product-service offerings, is worth the cost in their minds. Even economic ruin of the country is worth the price in their minds. Academic success is the only thing that matters to this faction since it is the only way to college, which is the only way to the "elite ruling class." "Damn the torpedoes, full steam ahead!" is the appropriate analogy for such blind foolishness.

This all began when, as the National Education Association (1918) criticized, colleges were unduly influencing secondary schools to focus attention on academic requirements for college entrance while neglecting the vast majority of the school age population. "[H]igher institutions of learning are not justified in maintaining entrance requirements and examinations of a character that handicap the secondary school in discharging its proper functions in a democracy." (pp. 19-20)

Carnoy et al. compare the purpose of elementary schooling, which better fit standards-based accountability reform policies, to high schools, which do not fit such policies. In particular, a key point is made with the following: "high schools are the place where major life decisions are made about students in their transition to adulthood and further education." (p. 197) Academic competencies that colleges demand are irrelevant to a majority of people – think about the way algebra is currently taught as just one example – and yet the demands of postsecondary institutions dictate what is required of everyone. If this is not seen as demented, there is no hope.

The authors make the comparison between elementary and secondary schooling:

> *The magnitude of change demanded from high schools – as organizations – is dramatically different from what is demanded of elementary schools.*
>
> The demands of standards-based reform call for a clarification and articulation of goals in elementary schools, for focused attention on instruction, for improvement and intensification of effort at what they were already designed to do – to prepare third graders to enter fourth grade, and fourth graders to enter fifth. High schools, on the other hand, are being asked to take on a new task – something they were not designed to do – to prepare students for a defined minimum academic standard, and to get all students to graduate by achieving that standard. We have certainly not organized high schools so that all students would take the same content, or meet the

same standards to graduate. In fact, comprehensive high schools were historically designed to do precisely the opposite; since the highly influential midcentury Conant report[19], their design imperative has been to serve democratic purposes and accommodate diverse student populations by creating a wide range of programs, and a differentiated curriculum.

That idea of differentiation, of the "bell curve" of abilities in high schools, has been central to the organizational design of the comprehensive high school, from the tracks that accommodate the large bulk of students in the general courses to the small number of students along the right tail in Advanced Placement classes.

…Teachers across our schools, like educators in national opinion polls, embraced the ideal of high standards for all students as means of reducing … inequities, both within and across schools, as "the right thing to do," as "making the system more fair." Many "applauded" the state for stepping in not only to set common standards, but also to establish common assessments and accountability that would force the issue of equity across schools.

Though teachers might applaud the ideal of common denominators and equal access, they did not equate this with narrowing the curriculum to provide one common experience, or to achieve a common outcome for all students. Instead, many worried that too much emphasis on the common core of academic subjects would betray the ideals of the comprehensive high school and the values of the school community: individual choices for students, preparation for diverse career paths, accommodation for different talents, opportunities for extracurricular activities, and room for social growth. In particular, educators frequently expressed concern that the emphasis on academic tests is undermining the value of vocational skills.

…Holding all students, within and across schools, to the same high academic standards without sacrificing the advantages of that diversity would be a radical reform indeed; a formidable challenge for both policymakers and high schools. (pp. 176-78)

Carnoy et al. next point out that students are the ones who, in the end, are held accountable and suffer when unable to meet the standards.

The stakes are higher in high schools.
In high-stakes accountability at the elementary and middle school level, schools are the principal objects of sanctions. But in high school, high stakes are also aimed at students. State accountability policies are moving to make diplomas contingent on externally set exams, or on demonstrated competencies in externally set performance reviews. … Ironically, in a system theoretically designed to benefit students, high school students may be the only people held directly accountable *as individuals* for achievement scores.

That also makes the stakes higher for accountability systems. … [T]he actual denial of a diploma is a consequence of a different kind, with political and legal implications, as well as financial and logistical ones.

[19] https://catalog.hathitrust.org/Record/001282593

> Teachers we interviewed in our study made a sharp distinction between standards, which they applaud, and the standardized high-stakes tests they deplore. Positive statements about raising standards were often followed with a pause, or a "but" – and a prediction of negative consequences, of rising failure rates, and increasing numbers of dropouts. Here practitioners depart from the sentiments and expectations of the policymakers, for they are much more worried that there will be what one calls "a lost generation" – students who have not been prepared to meet the standards but will not be able to graduate without passing the tests.
>
> …State policymakers are caught in a dilemma around how high to set the bar for high school standards: if too many students fail, they risk a "lost generation" and a loss of public support; if too many pass, they risk not being taken seriously. (p. 181)

This perspective of the risk of too many passing is an evil of the first order. All students should be competent in what is required to function in a free society. It demonstrates the callousness of academics and policymakers toward individual rights and interests. Many individuals are to be sacrificed on the altar of academic faith so that a portion might benefit by it. This is why it's so important to determine what is really necessary to learn to be a full participant in society, rather than what is really necessary for pursuing a postsecondary education.

A very important point made by the authors regarding the duties of elementary and middle school educators is offered with the following statement:

> Teachers describe the problem as the long-term absence of standards and accountability in the system – a system that has historically promoted students to high school without providing them with the education they need to succeed in academic work. (p. 183)

Part of the problem is that many students should not be prepared for academic work, but, rather, should be prepared for technical work. Academic work is in the realm of abstract theory, whereas technical work is a combination of application and abstract concepts. This is a far superior method of learning and it delves deeper and broader in the minds of students. Such instruction would require significant change in assessment methods.

Another problem the authors point out is the unpreparedness of teachers:

> In many cases, high school teachers are being asked to teach what they have not taught before, what they are not certified to teach, or, quite simply, what they do not have the capacity to provide. The 1998 Department of Education figures show that twenty-eight percent of math teachers and fifty-five percent of physics teachers have neither a major nor minor in their subject; another analysis estimates the figures at one in three math teachers, and one in four English teachers, who are teaching "out-of-field." Those shortages are more critical factors in the schools that are the very target of the reform, where an administrator asks "who is going to teach this stuff?" Each high school teacher who either cannot or does not provide sufficient "opportunities to learn" the new standards translates into approximately 150 students who don't receive them. Not only are teachers faced with new courses (like the humanities), but many teachers, after years of teaching general classes, suddenly find themselves teaching what had been the honors track. (p. 189)

Perhaps the standards-based accountability movement's attack – intentional or unintentional – on the principles of the comprehensive high school, provides the opportunity to shift the paradigm to the more successful models of Europe – in particular in Switzerland, Austria, Denmark, Finland, and Germany. However, there are some very good examples in France of craft schools, and crafts should not be forgotten since they have much to offer for the lower secondary years – if not as careers, certainly as foundational preparation for the technical world.

The Letter versus the Spirit of Education

The concept *the letter of the law versus the spirit of the law* provides a good analogy for the problem with the way tests are written. Also, the *essence* versus its *substance* also provides a nice reference point to consider the problem with our current testing regime.

To get the point across, let us look at two scenarios that require a different knowledge base. The building of a bridge requires one to follow the *letter of the law* since safety requires attention to detail. Obviously, engineers need extensive knowledge of math and science. However, a constitutional attorney will typically have no need for such precision but will need extensive knowledge in legal principles that are based on the essence of concepts. There is an important distinction between the *letter* and *spirit* of education demonstrated here, but the educational bureaucracy looks exclusively to the *letter* since tests are designed with only this in mind.

The National Academies of Sciences distinguishes memory abilities between younger and older people. The reasons for the differences are unimportant for the purposes of this paper. What is important is the difference between memory of specific details versus the "gist" or essence of something remembered. The National Academies provides:

> older adults are more likely to remember the moral of a story rather than its details and to report general rather than specific details of past autobiographical events. Studies show that declines in the specificity of memory likely begin in middle age, with increases in gist-based … memory already apparent by the time an adult is in his 50s.
>
> … The shift toward gist-based memory with age can lead older adults to be more likely than younger adults to remember the "big picture" or important implications. The shift toward pattern completion[20] also may enable older adults to note connections among events and to integrate across experiences, abilities that often are considered part of the wisdom that is acquired with age.

The point being made here is that specific memory abilities in younger people versus gist memory in older people is likely pointing to a fundamental flaw in the educational culture. Educators are interested primarily, if not entirely, with specific memory recall while ignoring gist memory. However, in the real world, it is gist memory that counts. One might compare this to the *spirit of the law* grounded in deep principles versus the *letter of the law* ground in facts and data. Therefore, educational programs should seek to help young people mimic the abilities of older adults by avoiding the current factual-data driven

[20] Joining a partial memory with other memories to reconstitute a complete memory representation.

testing regime, that is based on close transfer of learning, and in its place pursue the development of far transfer that develops trouble shooting and innovation abilities.

A Perspective on the Purpose of Grading

Mintz (Mar. 2016) asks the question "How can we make assessment more meaningful?" He argues for the importance of a rigorous assessment system, but then addresses the various objections to the current design.

> Many students complain that grading is arbitrary, inconsistent, and unfair, while many instructors grumble about grade inflation, the excessive amount of time devoted to grading, and the many complaints that grading prompts.
>
> Then there are the frustrations expressed by educational psychologists and psychometricians:
>
> - That a single, over-all grade conflates elements that need to be disentangled.
> - That **grades tend to overly reward lower-order thinking skills (such as memorization and recall) rather than higher order skills (involving analysis, application, and synthesis)**. (Emphasis added)
> - That grades too often fail to accurately reflect student learning or mastery.
> - That grades are frequently de-motivating and discouraging.
>
> … If we are to improve grading, we must first ask why we grade. Is it to rank students? To measure performance or knowledge or memory or higher order thinking skills? Is it to motivate students to study, or diagnose learning problems, or to assess mastery?

Mintz then lists the roles of grading: To inform students how they are progressing; to motivate students to master material; to diagnose; to evaluate levels of competencies; and to teach students how to be self-learners.

Mintz provides bullet points of what should be graded: Application of skills and knowledge, effort, participation, progress, and outcomes. These are general enough points that apply to any type of educational effort, be it academic or otherwise. He then points out what grades can be:

- Holistic, or targeted (based on discrete assignments);
- Norm referenced or criterion referenced (that is, grading can be relative to their classmates or based on predetermined criteria); and
- Calculated or judged based on subjective or objective criteria;
- We can provide opportunities for extra-credit or retakes and revisions – or not;
- Standards-based grading: Under this approach, students must demonstrate proficiency on well-defined course objectives;
- Achievement-based grading: Here, assessment is based on how far students go beyond minimal expectations;
- Mastery based grading: Students must retake assignments until an acceptable level of mastery is achieved;

- Specifications Grading: Create detailed “specifications” on what it means to adequately do an assignment and design assignments that give the students opportunities to demonstrate they have met these specifications;
- A game-based approach: This involves incentivizing attainment of certain learning goals by awarding points for completing certain assignments or badging certain accomplishments;
- A simplified approach: This involves replacing fine-grained assessments with more general categories, such as "Exemplary," "Accomplished," "Promising," and "Developing;"
- A proficiency-based approach: This approach focuses on progress toward clearly-defined learning objectives. Grades, from this perspective, communicate what a student is able to do, but gives them extended time to practice and develop their skills.
- Adopt a multi-tiered assessment strategy that assesses performance in varied ways. These might include checks for understanding, practice sets, project-based assessments, and team-based assessments. In other words, include application and creation of knowledge within your grading toolkit.

The Way We Retrieve Information

Willingham’s research (2007, pp. 192-93) offers insights into the way humans learn and is a must-read for anyone who teaches. This is a text on cognitive psychology and it provides foundational insight into the process of learning.

One insightful example is where Willingham references the term “chunking” and offers this: “We defined a chunk as a unit of knowledge with subcomponents that are related to one another, often semantically; because they often occur together, it is possible to think of these subcomponents as a single unit. The term *chunking* refers to the process of creating a chunk.” He then proceeds to provide examples of how chunking works. The first example is simple: He asks the reader to see if they can memorize the letters “FB ICB SNC AAP BS.” He then explains, “You’d remember some of the letters, perhaps all, but it would take some effort. Think how much easier it would be if I gave you exactly the same list of letters but I paused in different places. FBI CBS NCAA PBS. Both lists are organized into chunks by the pauses, but for the second list the chunks derive meaning from prior knowledge.”

Willingham then provides applied versions of chunking to demonstrate how it really works. He references a series of experiments using chess as the application to demonstrate chunking. Two groups of participants were selected: non-experts and experts. A chessboard was set up at a mid-game stage. Participants were given 5 seconds to observe the board and then were asked to duplicate the arrangement of chess pieces on another blank board. “Nonexperts got 8 or 10 of 32 pieces correct, but chess experts got nearly all of them right every time. Why? Chess experts have much more background knowledge. When they look at a chessboard, it looks to them like the second version of the letter task (FBI, CBS, NCAA, PBS). … These sorts of expert knowledge effects have been demonstrated for baseball, dance steps, bridge hands, maps, and music.”

This demonstrates how extensive a time frame is required to become highly competent. Therefore, it is extremely important to consider regarding the expectations we place upon students’ competencies. It becomes evident that the observance of details and the ability to connect them into chunks takes a

significant amount of time. Instead of expecting students to become experts, it makes more sense for them to become novices so they are simply familiar with a subject area until the need arises for them to become experts in a vocation or avocation. However, the two areas that do require expert abilities are language and math to the extent every citizen needs to fully participate in an economic and political life in a society. Anything beyond this must be dependent upon what their future careers will require of them.

Another characteristic that experts demonstrate that contradicts the way tests are conducted has to do with speed of answering questions. The National Research Council (2000, p. 49) offers the following: "[An] important characteristic of expertise is the ability to retrieve relevant knowledge in a manner that is relatively 'effortless.' This fluent retrieval does not mean that experts always accomplish tasks in less time than novices; often they take more time in order to fully understand a problem."

This contradicts the belief that speed of answering questions for tests is important and that it measures intelligence. This points out how dangerous current assessment tests can be. Many individuals are disenfranchised due to this idea. It's dangerous because it does GREAT harm to those who are not in the sprinting crowd, with government power and funding being used to harm the disenfranchised. Speed of answering questions may be one weakness of an individual and if this weakness effects grades that are on the borderline of being "college ready," it can make the difference between success and failure even though it has nothing to do with real world demands. Therefore, judging students based on how fast they get through tests is unacceptable.

GED

The GED may provide doors to be opened to many people who would otherwise forego an education. I remember two of my friends ending their high school education after sophomore year. During the summer of that year, they studied for and passed the GED test. That fall they were enrolled in the local community college and two years later had their associate's degree when all I had was a high school degree that offered nothing but the opportunity to say I did my time and was given the "permission" to attend college.

I think this is a perfect demonstration of wasted resources expended on a large percentage of the teenage population. If individuals are capable of passing the GED test, they are ready to move on to something better regardless of their age. To do otherwise may very well reflect the desire to trap individuals in the system in order to indoctrinate them (the less offensive term used is to "socialize" them) in the Progressive political philosophy. Of course there could be other reasons to stick with this rigid system, but the reins of control need to be handed over to students and their parents so **they** can decide what's best, especially if an individual has achieved literacy and numeracy. It should certainly not be in the hands of pompous and arrogant government bureaucrats who really don't have a clue.

Mortrude (2016) offers insight into what the GED can do to contribute to this noble cause that will certainly further purposes of achieving greater equity. First she points out that the GED Testing Service

> has lowered the passing score on its subject area high school equivalency tests from 150 to 145. This change … provides states a unique opportunity to raise the GED's profile in career pathway systems by aligning assessment with policies that award credit for prior learning. According to

> data in multiple states, GED-passers needed less college remediation than their high-school-graduate peers, meaning the test was gauging not just high school equivalency but also college preparedness.
>
> The GED Testing Service is now encouraging states to adopt a three-tiered range for passing scores, whereby a score of 145 indicates high school skills, 165 indicates college-ready skills, and 175 indicates college-level skills.

GED recommends that a score of 165 or higher should provide proof of postsecondary education readiness, so colleges should not require a placement exam, which would reduce waste in resources and individual's time. In addition, it is being suggested that if individual's scores are high enough – 175 or above – in subject areas, college credit should be awarded to them. This too would reduce waste in resources since once an individual has competency in a given subject, why should the same subject be studied yet again? This is a **serious** problem students currently labor under and is a major contributor to why many people forego higher education. Mortrude states, "GED students should have [the] … option to jump-start their postsecondary education while completing the high school equivalency."

Given the changes that are occurring in secondary education, it appears the current high school system is an outmoded program that will either need to dramatically change or it may face extinction.

Finland Provides a Very Good Example

Finland is a success story that all societies should analyze to get ideas of what would work in their countries. Hancock (2011) points out there are no mandated standardized tests in Finland, apart from one exam at the end of students' senior year in high school. There are no rankings, no comparisons or competitions between students, schools or regions.

> … The differences between weakest and strongest students are the smallest in the world, according to the most recent survey by the Organization for Economic Co-operation and Development (OECD).[21]
>
> … Ninety-three percent of Finns graduate from academic or vocational high schools, 17.5 percentage points higher than the United States, and 66 percent go on to higher education, the highest rate in the European Union. Yet Finland spends about 30 percent less per student than the United States.
>
> … "We prepare children to learn how to learn, not how to take a test," said Pasi Sahlberg, a former math and physics teacher who is now in Finland's Ministry of Education and Culture. "We are not much interested in PISA. It's not what we are about."

[21] Here we see an example of the leveling effect when education is taught properly. To make my point clear, I mean that instead of leveling being a downward trend, as it historically meant, it is an upward trend (eugenicists will hate this since they embrace the selectivity paradigm), whereby even those who will not hold positions of authority in business or government will be more cultured, well adjusted, more affluent, and be significant contributors to the community. This is an example of a healthier society to live in.

… Teachers in Finland spend fewer hours at school each day and spend less time in classrooms than American teachers. Teachers use the extra time to build curriculums and assess their students. Children spend far more time playing outside, even in the depths of winter. Homework is minimal. Compulsory schooling does not begin until age 7. "We have no hurry," said Louhivuori. "Children learn better when they are ready. Why stress them out?"

… Not until sixth grade will kids have the option to sit for a district-wide exam, and then only if the classroom teacher agrees to participate. Most do, out of curiosity. Results are not publicized. Finnish educators have a hard time understanding the United States' fascination with standardized tests. "Americans like all these bars and graphs and colored charts," Louhivuori teased, as he rummaged through his closet looking for past years' results. … "It's nonsense. We know much more about the children than these tests can tell us."

… Some of the more vocal conservative reformers in America have grown weary of the "We-Love-Finland crowd" or so-called Finnish Envy.[22] They argue that the United States has little to learn from a country of only 5.4 million people – 4 percent of them foreign born.[23] Yet the Finns seem to be onto something. Neighboring Norway, a country of similar size, embraces education policies similar to those in the United States. It employs standardized exams and teachers without master's degrees. And like America, Norway's PISA scores have been stalled in the middle ranges for the better part of a decade.

… Finland's expanded system of vocational … schools … are attended by 43 percent of Finnish high-school students, who prepare to work in restaurants, hospitals, construction sites and offices. "We help situate them in the right high school," said then deputy principal Anne Roselius. "We are interested in what will become of them in life."

Finland's schools were not always a wonder. Until the late 1960s, Finns were still emerging from the cocoon of Soviet influence. Most children left public school after six years. (The rest went to private schools, academic grammar schools or folk schools, which tended to be less rigorous.) Only the privileged or lucky got a quality education.

… In 1963, the Finnish Parliament made the bold decision to choose public education as its best shot at economic recovery. "I call this the Big Dream of Finnish education," said Sahlberg, whose upcoming book, *Finnish Lessons*, is scheduled for release in October. "It was simply the idea that every child would have a very good public school. If we want to be competitive, we need to educate everybody. It all came out of a need to survive."

Practically speaking – and Finns are nothing if not practical – the decision meant that goal would not be allowed to dissipate into rhetoric. Lawmakers landed on a deceptively simple plan that formed the foundation for everything to come. Public schools would be organized into one system of comprehensive schools, or *peruskoulu*, for ages 7 through 16. Teachers from all over

---

[22] When a faction doesn't have a good argument against a position, it inevitably turns to name-calling.

[23] This is a red herring. Finland can be compared to a State in the United States. And, after all, we do not have a national system of education, we have State run systems. Also the reference to 4% foreign born having anything to do with the issue demonstrates a form of bigotry. In essence it is saying that foreigners or people not of the white race are incapable of learning under a system like Finland's. This is nonsense.

> the nation contributed to a national curriculum that provided guidelines, not prescriptions. … The second critical decision came in 1979, when reformers required that every teacher earn a fifth-year master's degree in theory and practice at one of eight state universities—at state expense. From then on, teachers were effectively granted equal status with doctors and lawyers. … By the mid-1980s, a final set of initiatives shook the classrooms free from the last vestiges of top-down regulation. Control over policies shifted to town councils.[24] The national curriculum was distilled into broad guidelines. National math goals for grades one through nine, for example, were reduced to a neat ten pages. Sifting and sorting children into so-called ability groupings was eliminated. All children—clever or less so—were to be taught in the same classrooms, with lots of special teacher help available to make sure no child really would be left behind. The inspectorate closed its doors in the early '90s, turning accountability and inspection over to teachers and principals. "We have our own motivation to succeed because we love the work," said Louhivuori. "Our incentives come from inside."

College Admission Tests

Contrast Finland's desire to help every child, to an article written by Hoover (2016) about U.S. universities. Hoover interviewed professor Alexander Astin, founding director of UCLA's Higher Education Research Institute, about his book *Are You Smart Enough? How Colleges' Obsession with Smartness Shortchanges Students*. While Astin's analysis was focused on the university culture, it is equally applicable to our secondary education system in general. Fundamentally, Astin "believes that too many faculty members 'have come to value merely *being* smart more than *developing* smartness.'"

Astin reflects on the university culture where "he finds more concern with 'acquiring' smart students, as defined by conventional metrics, than with helping students improve after they enroll. … 'If colleges were instead to be judged on what they *added* to each student's talents and capacities, then applicants at every level of academic preparation might be equally valued.'"

Academics expend all their energy on the "smartest" students – as measured by test scores – at the expense of all others. "We have created an institutional structure that reflects this bias. Teaching an average student doesn't get any value in academia." Keep this in mind next time the universities in your State ask for more funding.

In Hoover's interview, Astin states,

> I'm a psychologist, trained in statistics. The people who make these tests are like me, kind of mesmerized by the normal curve and the elegance of the stats that underlie the normal curve. We unthinkingly come up with scores that inevitably rank students. But there's not much information in an ACT or SAT score. … These very narrow measures of standardized tests and grades, when they're used competitively to sort and select rather than to educate well our students, they put large segments of society at a tremendous competitive disadvantage.
>
> In the domains of creativity, especially artistic creativity, there's very little overlap with traditional smartness. Leadership is another fascinating area. So many college mission statements

[24] U.S. school systems had outstanding programs before the Horace Mann, German influenced centralized bureaucracies took control in the 19th century. Finland's success reveals we should return to this model.

> talk about leadership, and it has very little relationship to SAT smartness. Then there's what we might call character – honesty, trustworthiness, authenticity.

Astin points out how the academic culture immerses faculty in an environment that worships "smartness" and how academics often try "to show off our smartness to each other, or avoid being judged as not smart enough." It sounds as though the culture is filled with very insecure people attempting to erect a façade of intellectualism in order to be part of the "elite" clique. This is why I consider them *academics* rather than *scholars*. There's a **big** difference between the two. The former is all about standing "above" and showing off, while the latter has no thought of such matters (think of Einstein) and is only interested in furthering understanding, which we might categorize as developing greater wisdom. Academics are of a common mindset, whereas scholars are a rare breed.

I believe Astin's analysis offers insight into the academic culture that should give citizens pause as it relates to the trust they have in academics. Academics, in contrast to scholars, are not the loving and compassionate people they make themselves out to be. They are some of the most selfish and egotistical people one can find in our culture and they need to be brought back down to earth where they belong. In addition, they are subject specialists, with few of them being able to transfer their area of understanding to any other area of society, which makes them no more or less informed than the average person outside their subject specialty.

Johnson (2016) reports on a survey that supports Astin's position on the marginal information one can glean from college admissions tests.

> An incoming college freshman's high school GPA provides more insight into first-year academic success than popular college admissions tests such as the SAT and ACT, according to a National Association of College Admissions Counseling survey of 400 colleges and universities.
>
> The survey also found that slightly more than half (51 percent) of these schools conduct what are known as predictive validity tests – reviews to determine how closely their various admissions criteria such as testing, class rank, recommendations and writing ability correlate to college success. "Overall, it is clear that high school grades are by far the most significant predictor of college academic achievement," the NACAC authors conclude in the study.
>
> However, responses from participating schools also revealed that the SAT and ACT and other tests still made "a significant contribution to the ability to predict college academic performance."
>
> The authors advise caution in drawing broad conclusions on college success, noting that there is little uniformity in the way each school conducts its review process.

Johnson quotes Michelle Hernandez, an author on the subject of college admissions: "The majority of students applying to elite colleges spend hundreds of hours doing SAT/ACT prep when they could be pursuing scholarly activities. Many New York City families will spend over $20,000 on SAT prep…."

Osborn (2016) provides the following:

> Susan Gershenfeld and co-author Denise Ward Hood looked at 1,900 University of Illinois students who graduated within six years of enrolling. They found that ACT scores of graduates and those who dropped out were almost identical. 69% of those who started college finished during the time frame, with an average ACT of 24.5. The students who started with this same cohort but didn't complete their degrees had an ACT of 24.1.

Kiener et al. (2015) point to the injustices that tests such as ACT and SAT have produced:

> [I]n less than a century, we have debunked the idea that the ACT or SAT measure intelligence. We have also seen the multibillion-dollar test preparation industry boom, with wealthy, college-driven families having much greater access to the test-taking strategies necessary to earn Ivy-level scores. The ACT and SAT have essentially become better indicators of family wealth than college readiness, closing doors on the very kids they were originally intended to help. What's more, we have also debunked the idea that the ACT and SAT can predict if students are ready for college.

Jaschik (2016) reports that ACT scores were down in 2016, which is attributable to the increase in the percentage of high schoolers taking the test.

> ACT data show that 64% of high school seniors in the class of 2016 took the ACT this year, up from 59% last year and 52% in 2012. … Generally, when a larger share of students take a test … scores go down. Score drops were the largest in states that have just started to require all students to take the ACT.

It must be taken into account that there is somewhere between a 25% to a 30% high school "drop-out" rate (those who pass the GED test cannot be counted as "high school graduates" even though States do include them in their figures as such) so the 64% in the 2016 cohort mentioned above represents approximately 45% to 48% of the entire age group in 2016.

> ACT has benchmarks for predicting college success for each section of the test. These too are showing declines. This year, 38 percent of test-takers met the benchmarks in at least three of the four subject areas tested (English, math, reading and science), which according to ACT shows that they have 'strong readiness for college course work.' That's down from 40 percent in 2015. The percentage of test-takers who did not meet any of the benchmarks increased to 34 percent from 31 percent.

Again, based on a high school graduation rate of 70% to 75%, the 38% who have a "strong readiness for college course work" is actually only 26.6% to 28.5%. This correlates well with the real needs of the economy – 28% of jobs in the economy require a college degree – as reported by The Bureau of Labor Statistics.

The drop in "college readiness" shows that not everyone is gifted in taking tests, nor in studying abstract and disconnected data that is sold to the public as an education. This trend will either not change or improve ever so slightly, regardless of public investment, demonstrating that our educational system is beating a dead horse. Therefore it's time to change the paradigm!

Bidwell (2014) provides another view:

> More students than ever are taking the ACT college admissions test, but student achievement remains flat, as nearly one-third of students are not meeting any college readiness benchmarks, according to a new report from the testing agency.
>
> More than half of students who graduated high school in 2014 took the ACT, the company finds in its annual report, "The Condition of College & Career Readiness"…. But of the 1.8 million test-takers, just 26 percent met college readiness benchmarks in all four subjects: English, reading, mathematics and science. In total, 47 percent met no more than one college readiness benchmark.
>
> …Overall, 64 percent of test-takers proved they were college-ready in English, 44 percent in reading, 43 percent in math and 37 percent in science.
>
> …Student achievement in science has generally trended upward, to 37 percent of students meeting the benchmark in 2014 from 29 percent in 2010. But all other subjects have been mostly flat in the last five years, or have declined, the report shows. The percentage of students meeting the reading benchmark has dropped 8 percentage points since 2010, when 52 percent of test-takers met the goal.

Schaeffer (2015) points out that a growing number of colleges and universities are reducing or eliminating ACT and/or SAT test results. "More than 850 accredited, bachelor-degree granting schools now do not require all or many applicants to submit SAT or ACT scores." This trend is expected to continue given many researchers' findings.[25]

Koretz et al. (2014) provide a report critical of high-stakes testing. Their findings

> raise serious concerns about the effects of high-stakes testing on instruction. The past several years have seen continuing debates about appropriate and inappropriate teaching to the test. Skeptics about test-based accountability, including several of us, have suggested that undesirable narrowing of instruction is one likely consequence of high-stakes testing. Supporters of test-based accountability, on the other hand, argue that focusing on the content of the test is desirable, as long as test-based accountability leads teachers to focus on broad areas of knowledge and skills measured by the test rather than on content specific to the test question. Our results seem clear enough: to a substantial degree, teachers in this district must be focusing on content that is specific to the particular test used for accountability, rather than trying to improve achievement in the broader sense that we would all desire. (pp. 16-17)

It can be imagined that teachers who teach based on broad areas of knowledge and skills will be criticized if their students don't perform as well on these tests as those who teach students content specific to the tests. The tendency will be irresistible to teach to the tests.

---

[25] See Hiss, W. C. and Franks, V. W., *Defining Promise: Optional Standardized Testing Policies in American College and University Admissions*, Feb. 5, 2014. http://offices.depaul.edu/enrollment-management-marketing/test-optional/Documents/HISSDefiningPromise.pdf

Lynch (2016) criticizes standardized assessment as causing more harm than good.

> A number of problematic patterns emerge when severe consequences result from failure of students to perform at required levels on assessments. When a new test mechanism is introduced, teachers and students are not familiar with the content or the format of the test. Scores in this environment tend to drop. As time passes, and the teachers and students become acquainted with the test and the subjects involved, the scores climb until they reach a plateau. After that point, additional improvements in the scores become less achievable. I believe that there are three central reasons why we should do away with standardized exams altogether. Here they are:
>
> 1. It encourages teaching to the test.
>
> Because of the stress and the potential ramifications of poor scores, teachers began teaching to the test. More time was spent with academically challenged students, sometimes even causing them to miss other classes, and the amount of time spent exclusively on material covered by the test was greatly increased. Teachers sometimes changed their teaching styles to accommodate content included on tests. Their students could demonstrate rote knowledge but could not apply it or appreciate it in context. Test-score improvement could be attributed to the alteration of the teaching parameters, but this would not improve students' problem-solving abilities or critical thinking skills.
>
> 2. It sacrifices means for the end.
>
> Another issue concerns ethical pedagogical methods. In some schools, limited resources are allocated to students who are not likely to make the largest impact on the school's annual yearly progress. Teaching time is reduced with students who will obviously complete upcoming assessments successfully, as well as with students who appear to have a very slim chance of reaching proficiency levels on the assessments. Instead, teachers focus their attention and efforts on the students who are close to the "cutoff line." Clearly, these students could have the greatest short-term impact on the school's standing; however, do the practices represent a moral approach to education?
>
> 3. Its power in education is too broad.
>
> The pressure to achieve the necessary test scores permeates all facets of the school system. Superintendents and principals have a lot riding on these results. Their positions may depend on the success of their students on required assessments. If test scores begin to drop, out go the administrators, regardless of the true cause. This pressure to achieve on standardized assessments trickles down to teachers, and subsequently to the students.

This is a common dilemma with too much government intervention and/or overregulation. It skews and twists decisions and the allocation of resources that would otherwise be utilized in far more productive ways. Of course, the argument for so much control is that without a tremendous amount of government oversight, we could not achieve perfection or that people cannot be trusted (as if government bureaucracies had god-like powers and were immune from human evils). There is no such thing as

perfection and bureaucracies are more dangerous than autonomous communities. The complete failure of the current system offers all the evidence one needs to see the truth in this statement.

Changes in Remedial Education Demonstrate Serious Flaws in College Entrance Exams and Expectations of Prospective Students

Lubbers (2016) defines a new trend in the wind: *corequisite remediation*:

> To achieve a better result for Hoosiers, Indiana has emerged as a national leader in embracing an approach that treats remediation as a "corequisite" rather than a "prerequisite" to student success. Instead of a long series of non-credit, remedial courses, corequisite remediation enrolls students directly in college-level courses with extra support that gets them up to speed while working toward their degree.
>
> The results are encouraging. At Ivy Tech Community College, corequisite courses in math and English have nearly doubled student pass rates in those subjects. Before corequisite remediation only 29% of students completed their introductory math courses in two years; now 64% complete the course in one year. In English, 55% of students completed their introductory course in one year, compared to 37% completing in two years before corequisite remediation.
>
> Indiana's success at scaling corequisite remediation statewide at Ivy Tech is receiving national notice. A new report by Complete College America – called "Spanning the Divide" – spotlights the impressive progress in Indiana along with a handful of other leading states that are seeing similar results.
>
> Complete College America's report also shows that successful corequisite remediation programs target the level of support students need to be successful and ensure that math classes students are required to take are aligned with their degree program and career path.

Bethke (2016) offers another summary of this report:

> Traditional prerequisite remediation is failing, according to a new report from Complete College America; but a new form, called corequisite remediation, may be able to provide triple the success rates in a quarter of the time for unprepared students, allowing them a better chance at eventually graduating.
>
> The report, titled *Corequisite Remediation: Spanning the Completion Divide – Breakthrough Results Fulfilling the Promise of College Access for Underprepared Students*, is the first of its kind on the topic and is specifically optimized for online viewing and interactivity. It highlights what it says are dramatic results from statewide efforts in Colorado, Georgia, Indiana, Tennessee and West Virginia, and also allows users to explore remediation numbers in their own state.

> As described in the report, 42 percent of all U.S. students, which amounts to more than 1.5 million students annually, begin college in remediation (prerequisite coursework that does not count toward a degree, usually costing students and states time and money).
>
> … With corequisite remediation, however, students can enroll directly into college-level courses and receive academic support alongside their regular classes, states the report. Instead of taking numerous prerequisite non-credit courses, students learn the needed material while working toward their degree. Within their college-level courses, additional class periods or customized support in a lab provide opportunities for academic support and tutoring when bridges need to be gaped.
>
> While the national average of students enrolled in remediation who complete their gateway courses in math and English over 2 years is only 22 percent for each, the numbers soar to 55 to 70 percent in states like Georgia, Indiana, Tennessee, West Virginia and Colorado that have adopted corequisite remediation.
>
> Moreover, these higher numbers are achieved in dramatically less time – usually only one year or semester.

Smith (2016) cited another report that revealed the same trend regarding remedial education.

> A student placed in remedial math has a better chance of succeeding in college by taking college-level statistical courses with additional support instead of developmental math, according to a new report published today in the American Educational Research Association's *Evaluation and Policy Analysis* journal.
>
> The research -- conducted by researchers at the City University of New York – found that for first-year college students who tested as needing remedial algebra, 56 percent of those assigned to and enrolled in college-level statistics passed, while only 39 percent of those assigned to and enrolled in elementary algebra passed. The students who passed the statistics course also subsequently accumulated more credits.
>
> "The majority fail the remedial elementary algebra. They don't like it and they don't like taking it," said Alexandra Logue, one of the study's authors…. "But if they take statistics, they have a more positive experience and they're much more likely to pass…."
>
> "Having students take a math course that's most useful to them in their major just makes sense," said Robin Ozz, president of the National Association for Developmental Education. "The statistics is in the quantitative reasoning path…. They're still getting that logical deduction and critical thinking just in ways they can use in their majors. There are a lot of ways we can teach that besides putting it in algebra."
>
> Math remediation can be a huge barrier to increasing graduation rates. More than 80 percent of students who are assessed as needing math remediation never complete the course work, according to the report.

This demonstrates that the problem lies primarily with the courses and assessment tests, not the students. Algebra courses in particular are notorious for how poorly designed and disconnected from the real world the curriculum is. I've never met one person who has said that algebra classes are well designed; and for those who have had no use for algebra in their lives, not one has said algebra benefited them. They only have bad memories of it. So the faith-like belief that algebra provides logical deduction and critical thinking type of mental development is a fallacy due to how poorly the curriculum is designed.

The findings on corequisite successes also demonstrate how monopolies behave: Since there is no competition, those delivering the product/service have little incentive to change, and the levels of failure observed can be blamed on the customer without any repercussions. It's good to see academia finally doing something to improve the situation at the postsecondary level. It must have taken a tremendous amount of outside pressure for this to have occurred.

The Abstract of the paper Smith reported on (Logue et al., 2016) states:

> Many college students never take, or do not pass, required remedial mathematics courses theorized to increase college-level performance. Some colleges and states are therefore instituting policies allowing students to take college-level courses without first taking remedial courses. … We randomly assigned 907 students to (a) remedial elementary algebra, (b) that course with workshops, or (c) college-level statistics with workshops (corequisite remediation). Students assigned to statistics passed at a rate 16 percentage points higher than those assigned to algebra, and subsequently accumulated more credits. A majority of enrolled statistics students passed. Policies allowing students to take college-level instead of remedial quantitative courses can increase student success.

Page 1 starts with:

> Colleges in the United States assess a total of about 60% of their new freshmen as unprepared for college-level work, most often in mathematics. College policies usually require such students to complete remedial courses prior to taking college level courses in the remedial courses' disciplines, based on the purported theory that students need to pass the remedial courses to be able to pass the college-level courses. However, the percentage of students successfully completing remedial courses is low.

Analyzing Tennessee's work with corequisite efforts, Belfield et al. (2016) found "gains in cost-effectiveness from moving from prerequisite to corequisite remediation under almost all plausible scenarios. Based on these Tennessee data, the success rates from corequisite remediation indicate a more efficient instructional system for students who enter college academically underprepared."

Devaney (2016) reports on Illinois' efforts in implementing corequisite remediation: "Nearly half of the state's 48 community colleges have implemented co-requisite remedial courses in English, math, or reading that place college students into remedial and college-level courses in the same subject at the same time. These co-requisite courses allow students to receive targeted support to help boost their understanding and learning of the college-level material."

Palmer (2016) reports, "… the Colorado Community College System in 2012 decided to redesign remediation. The redesign was extensive but one of its key features was the introduction of co-requisite remediation."

Reporting on Florida's work in this area, Travis (2016) offers:

> South Florida's less prepared college students are skipping remedial education and heading straight to college-level courses, with better results than many expected. Community college students who scored low on a placement test used to be required to take noncredit remedial classes. But the state in 2014 allowed most students to decide for themselves whether they needed these classes. Since then, enrollment in remedial classes has dropped 41% at Palm Beach State College, 69% in Broward College. Most of these students instead are taking regular college math and English classes during their first year. Many college officials feared high failure rates as more students took classes before they were ready. But so far, that doesn't seem to be the case. The percentage of students passing college-level classes hasn't dropped dramatically, and in at least one case, the rate improved.

This demonstrates how poorly designed college entrance exams are.

Ashford (2017) references the successes in New Jersey and Michigan in reducing the number enrolling in remedial courses. She mentions how Warren County Community College eliminated placement testing with the results that "the college's graduation rate went from about 20% to 40%.... In the past, a student placed into a developmental class at WCCC had just a 7% chance of eventually graduating. Among students who were in remedial classes in both English and math, no one graduated. [It is believed] students will succeed, even if they are not considered college ready…."

Finally, we see real effort to help students through the system rather than weed out those who have not historically fit the academically optimized system.

This section of the essay might deviate somewhat from the course the Applied Education Foundation is heading down, but the purpose of its inclusion is to demonstrate that current assessment testing regimes – such as ACT, SAT, and college entrance exams to determine "college readiness" – are misguided and dangerous. Conclusions gleaned from them have done tremendous harm to untold millions of individuals for over a century by blocking access to useful, necessary, and important credentials that contribute to full participation in local communities in particular and society in general.

The Investment in Math and Science Rather Than Career Education

Lemann (1995) points to an important historic turn that may have determined the inferior place career education has been relegated to in our educational establishment in sharp contrast to the prominent place it holds in certain European nations. The elitism assessment tests were designed to establish – originally devised for eugenic purposes –, focused most, if not all, of society's educational resources in this direction.

Most people historically and currently want to be part of the elite, which is due, for the most part, to their insecurities and fears that dominate their existence – that is, they don't want to appear as "inferior" to others. They acquire their sense of self-worth from the good opinion of others rather than from within. This is due to their lack of confidence and therefore they must attempt to compensate their feelings of inadequacy through accomplishments they can boast about, which is one reason credentialism has grown exponentially since higher education has come to dominate our society.

Lemann reveals what WWII and the Korean War did for education. WWII allowed for the launching of the mass-use of assessment tests in order to "qualify" candidates for officers' candidate school. The Korean War allowed for the use of assessment tests in order to defer from the draft "intelligent" individuals (working class type of jobs were not considered worthy endeavors hence the reason this sector was considered expendable which conforms with eugenic efforts). Self-preservation would have caused many to seek a way out of the draft (especially since it was not our war) and once the rationale behind it had been accepted, the reasons for it were forgotten in short order and people simply came to believe that assessment tests were simply part of the educational landscape. People forgot the reason for its popularity during this period, which, in many cases, was to avoid serving in the military rather than really desiring to go to college. This practice continued when President Johnson's obsessive and unjustified "police action" in Vietnam was raging. The outstanding economist, Milton Friedman, even commented that many university professors in the 1980s were of low quality because they ended up pursuing a Ph.D., and eventually becoming professors, as a means to dodge the draft. He said many were not of the scholarly caliber to be professor material.

Lemann points out how we were concerned about the lack of recruitable mathematicians and scientists by the time the Korean conflict was underway since American universities were not graduating enough of them to fill the demand. Many had previously come to the U.S. to escape Nazi Germany. When the war was over, German scientists stopped immigrating and therefore this pipeline of high-quality scientists declined precipitously.

It is interesting that this same argument was not used as it relates to craftsmen and technicians. The U.S. was built upon the abilities of European immigrants who were highly skilled, having been thoroughly trained in their home countries. However, when the flow of these immigrants subsided, just like the immigration of scientists, there was no educational system in place to replace such talents and abilities. Unfortunately, unlike the additional resources the U.S. started investing in math and science, little to no additional resources were allocated to career education.

The elitist assessment culture combined with Federal and State funds being made available to those who did not have the economic wherewithal to attend college, shifted educational goals away from the needs of most Americans and focused it on the opportunity to be part of the "elite" academic community with most resources being expended in scientific and mathematical directions. The inequity of this situation is readily apparent and using the majority of public educational resources for a select group of individuals is unjust and therefore illegal. This situation **must** be remedied through the development of a large variety of programs suitable for the many walks of life people will pursue.

Assessing the Educational System

I believe that which is frequently overlooked, besides a poorly designed educational system, as it relates to student performance is teacher, school, and district performance. As long as the established system can show an average bell curve, all is considered to be in perfect running order.

Wieman (2015) looked at defining quality from the perspective of teacher quality when he stated, "I propose a simple and operational definition: the effectiveness with which the teacher is producing the desired learning outcomes for the given student population. There are a number of outcomes we want for students, but at the level of the individual course or instructor, the degree to which students master the material and complete the course are primary." He was analyzing teacher quality as it relates to college instructors, but I wish to take his perspective and apply it to public schools.

Besides the very structure or paradigm of the current system itself, the bottom line comes down to how well teachers, schools, and districts perform, which can be measured by outcomes – and not how many students are "college ready." I would argue that if more than approximately 30% of a high school's graduating class applied to a college to earn a degree (as opposed to earning a certificate or for continuing education), the high school has failed its community. The reason for this is that the economy requires a mere 28% of the labor force to possess a college degree – associate's through graduate levels – as reported by the Bureau of Labor Statistics. Therefore, if a school or district is pushing the "college for all" mindset, they are harming many of those who do go to college, some of who will graduate and some will not. The 28% will benefit while the rest will be harmed due to college debt, underemployment, and the attendant loss in one's sense of confidence for having failed to benefit economically from the degree.

Hechinger (2015) points to the number of bachelor's degree holders needing to attend a community college to get training for a career that the university did not provide.

> A surprising one out of every 14 of the people who attend community colleges – widely regarded as low-tuition options for the less-well-prepared – has already earned a bachelor's degree, according to the American Association of Community Colleges. That's 770,000 students. At some community colleges, the proportion is as high as 1 in 5.
>
> …The phenomenon also has exposed a failure by some four-year universities to prepare their graduates for the kinds of jobs available in their surrounding regions, a longtime focus of community colleges, experts and observers say.
>
> "There's a lot of disciplines universities aren't offering," says Cecilia Rios-Aguilar, an associate professor of education at UCLA and director of its Higher Education Research Institute, who says it's no surprise that college graduates who want more specialized training or career changes are turning to community colleges. "The universities aren't keeping up."
>
> By definition, community colleges are more responsive to the needs of local employers than some universities, says Davis Jenkins, a researcher at the Community College Research Center.
>
> "Certainly the regional universities should be more customer-responsive," he says. The fact that they're not has driven up the proportion of students with bachelor's degrees at some community

> colleges. At one, Foothill College in California's Silicon Valley, the number rose to about 30 percent after the economic downturn before falling to about 19 percent now.
>
> Scholars have conducted little research into this group, whose numbers have surprised even seasoned educators.

This is yet another example of how out of touch colleges and universities tend to be.

Martin and Gillen (2011) address the need to assess educational institutions. They state:

> A popular notion within the academy is that teaching quality cannot be measured, but this is an article of faith, not a demonstrated fact.[26] Very few institutions have made a systematic effort to measure teaching quality, largely because the faculty is opposed to it and administrators have little incentive to discover true teaching value added. Faculty view their conduct in the classroom as beyond judgment, while for deans, knowing how serious some teaching problems are is a kind of trap: this obligates them to fix those problems in an environment where very little can be done. Further, if some professors are identified as truly exceptional teachers, their peers may resent it and the exceptional teachers may expect higher compensation in return. So, most administrators choose to leave that sleeping dog alone.
>
> One consequence is that colleges and universities scrupulously avoid competing on the basis of teaching metrics; choosing instead to compete on the basis of things that signal or imply quality, such as scholarly research, elaborate facilities, stately campuses, athletic teams, and extravagant entertainment. This competition accounts for most of the excess cost of college and for the decline in teaching quality.
>
> … There are serious inconsistencies and misleading assertions employed by those who make the claim that teaching quality simply can't be measured. It is more likely that, as faculty members, we simply do not want to be graded, even though we are hired to grade others.
>
> The first inconsistency in the argument against quality measurement is that if quality cannot be measured then grading is a fraud. When we measure classroom performance, we evaluate a variety of subjective aspects of student performance in order to assign a final grade.
>
> Another inconsistency is that we accept the measurement of scholarship quality but reject the measurement of teaching quality. Scholarship is not an objective concept. What constitutes scholarship varies by discipline and is very hard to evaluate across disciplines; things that are regarded as valuable contributions by one discipline seem not to be in others. Further, scholarship is not obvious even within a single discipline. Consider the endless discussions that go on within research departments about what journals are in the "top twenty-five," how to evaluate books versus journal articles, and how to adjust for multiple authors. Further, how do we evaluate citations, do all citations count regardless of the source? Suppose the article stimulated a number of citations because it contained a serious error?

[26] I would argue it's a means of escaping evaluation and responsibility. After all, who likes to be assessed? If you're in a position of authority, you are more than likely going to justify the reasons you should not be evaluated.

Notice that even though there is great subjectivity in evaluating scholarship, the information is sufficient to establish a brisk market for scholars. The fact that there is no comparable market for teachers is due to the fact that we refuse to put the same effort into measuring teaching quality as we do in measuring research quality.

Another misleading argument is the use of perfection as a club to fend off attempts to measure value added. For people who argue this way, any proposed measure of value added has to be perfect or it is rejected; they make perfection the enemy of the good. In other words, they reject what we already know about scholarship (it does not have to be perfect to serve us quite well) in order to forestall the inevitable accountability that will come with measurement of teaching value added.

A fourth misleading position occurs when someone argues that we cannot measure value added because it is impossible to predict how an individual student will perform; it is said there are too many random variables that influence an individual student's performance.[27] For example, two students with the same background may differ radically with respect to their motivation, their maturity, or their health. Technically, it is true we cannot predict individual student behavior. Fortunately, we do not have to predict individual behavior in order to measure either an individual professor's value added or an institution's value added.

… It is hard to overstate how much our refusal to measure value added costs society.

Further, the refusal to measure value added is very costly to members of the academy as well. Most of us know the incentive system is badly skewed towards research and this comes at the expense of teaching. The continuous widening in the income distribution over the last three decades and the stagnation in median incomes are, in part, the result of the high cost and declining quality of undergraduate education. Our ability to compete in an international economy is adversely impacted by the likely decline in teaching value added. The aggressive "mission creep" among institutions occurs because all the rewards go to research rather than teaching. The nations' most gifted would-be teachers choose not to become professors because the pure teaching track in higher education offers very few rewards.

…More information on all of these is available at the **National Institute of Learning Outcomes Assessment website**. Unfortunately, participation and reporting for each of these measures is generally voluntary, and it is interesting to note that those institutions at the top and bottom of the hierarchy typically refuse to participate.

These and other metrics will never be perfect, but they will be efficient; they will be enough to establish a market for world-class senior teachers and that will be a game changer for higher education.

Abdul-Alim (2016) also argues for the assessment of the educational establishment. The author quotes Fredrik deBoer of Purdue University:

---

[27] Yet our educational system, for the most part, is designed as a one-size-fits-all program that does not take into account these differences.

> Colleges and universities should embrace assessments of student learning in order to prove their worth as college costs rise…. "There's no enterprise in the world where we pay hundreds of millions of dollars and no one asks to see how well we're doing, except the defense industry."

Next, we need to look at what Nelson (2014) suggests. Nelson is the President of St. John's College, which is unique in what it teaches (the Great Books of the Western tradition amongst other subjects), how it teaches, and how it assesses. It is very traditional and truly a scholarly liberal arts college, which I have the greatest respect for! It is a college whose students learn how to think, which is extremely rare today!

Nelson takes the position that "trying to determine whether an institution effectively delivers knowledge to its students … downplays the role of students in their own education, placing far too much responsibility on teachers and institutions…." No doubt, students **must** work hard at their studies if they hope to be successful. However, it is a two-way street: If educators and their institutions have strategies and methods that are poorly designed and delivered, even the best students will struggle. Just because a teacher passes through an educational program and is awarded a credential, doesn't mean that person will be effective at teaching the subject. Therefore, both the students and the educational establishment **must** be assessed to see how they are performing.

Nelson does, however, cause one to ask questions and seek answers as to why he and so many other educators take issue with being assessed. Could it be that the system is so poorly designed that no one can expect educators to achieve high marks if assessed? What they may be saying without awareness of it is this: "What is taught and how it is taught is so terribly organized (which was simply handed down to us) that only a small percentage of the population of students can be expected to perform within such a bad system." Those who come from higher socioeconomic backgrounds have the advantage because their parents have the resources to invest in helping them bring some order out of the chaos. This provides educators cover so they can claim they are being highly effective in teaching since they have a percentage of students who demonstrate sufficient, though shallow, understanding of educational materials.

I propose this is the primary dynamic playing out in the educational establishment, in particular at the postsecondary level – since professors are not taught how to teach – and it helps explain why we are so stuck in a failed system.

Selingo (2015) analyzes how students assess themselves and how employers assess college graduates. A study released the month of this article found a

> disconnect between what employers need and the readiness of college seniors. In a pair of surveys by the Association of American Colleges & Universities, would-be graduates said college armed them with the skills needed for the job market. But employers disagreed. On a range of nearly 20 skills, employers consistently rated students much lower than they judged themselves. While 57 percent of students said they were creative and innovative, for example, just 25 percent of employers agreed.

Those within the academic community see themselves in a very different light than those outside of it. This uncovers the need for objective assessments being routinely utilized to determine the value, if any,

education curricula and programs provide individuals and society. This must be done from outside academia since vested interests will attempt to thwart the effort. Those who attempt to stop the effort should be barred from teaching.

Multiple Assessment Methods

Conley (2014) frequently references the common term "college and career ready," however, he admits, "we are only just beginning to understand what 'readiness' really means. In fact, the most familiar measures of readiness – such as grades and test scores – tend to do a very poor job of predicting how individuals will fare in their lives after high school."

An important idea that needs to be embraced is that college readiness and career readiness are two separate issues that need their own identities to be defined. If this is not understood, little progress can be expected since the academic track will overshadow career interests, which history has proven since 1894. The confusion that surrounds these two issues had its origins in The National Educational Association's 1894 report: *Report of the Committee of Ten on Secondary School Studies*. The Committee of Ten asserted that preparation for life and preparation for college were identical goals that can be achieved with the same educational program, in spite of the arguments to the contrary even at that time. Many railed against the Committee's position, but to no avail. This mistaken notion has yet to be resolved.

Conley starts off the report with a historical analysis, pointing out that American educators have been more concerned with repeatability of testing so that information will be consistent over extended periods of time rather than valid – "valid" meaning reporting what really matters in assessing the accomplishments and competencies of students.

Next Conley identifies the focus of assessment tests being on

> discrete questions, each one pegged to a very particular skill or bit of knowledge…. This focus on particulars has had a clear impact on instruction. In order to prepare students to do well on such tests, schools have treated literacy and numeracy as a collection of distinct, discrete pieces to be mastered, with little attention to students' ability to put those pieces together or to apply them to other subject areas or real-world problems.
>
> … What if understanding the parts and pieces is not the same as getting the big picture that tells whether students can apply knowledge, and, perhaps most important, can transfer knowledge and skills from one context to an entirely new situation or different subject area? If it's not possible to do these critical things, then current tests will judge students to be well educated when, in practice, they cannot use what they have been taught to solve problems in the subject area (what is known as 'near transfer') or to problems in novel contexts and new areas (known as 'far transfer') (pp. 2-3)
>
> [T]he longstanding American preoccupation with breaking subject-area knowledge down into small bits, testing students' mastery of each one, and then teaching those bits sequentially, may in fact be counterproductive. Rather than ensuring that students learn systematically, piece by

> piece, this approach could easily deny them critical opportunities to get the big picture and to figure out which information and concepts are most important.
>
> When confronted by a torrent of bits and pieces presented one after the other, without a chance to form strong links among them, the brain tends to forget some, connect others in unintended ways, experience gaps in sequencing, and miss whatever larger purpose and meaning might have been intended. Likewise, when tests are designed to measure students' mastery of discrete bits, they provide few useful insights into students' conceptual understanding or their knowledge of how any particular piece of information relates to the larger whole.
>
> The net result is that students struggle to retain information. Having received few cues about the relative importance of the given content, and having few opportunities to fit it into a larger framework, it's no wonder that they often forget much of what they have learned, from one year to the next, or that even though they can answer detailed questions about a topic, they struggle to demonstrate understanding of the larger relevance or meaning of the material. Indeed, this is one possible explanation for why scores at the high school level on tests … have flatlined over the past two decades, a period when the emphasis on basic skills increased dramatically.
>
> Ideally, secondary-level instruction guides students through learning progressions that build in complexity over time, moving toward larger and more integrated structures of knowledge. Rather than being taught skills and facts in isolation, high school students should be deepening their mastery of key concepts and skills they were taught in earlier grades, learning to apply and extend that foundational knowledge to new topics, subjects, problems, tasks, and challenges.
>
> And in order to provide this sort of instruction, teachers require tests and tools that allow them to assess far more than just the ability to recall bits and pieces of content. What is needed, rather, are opportunities for students to demonstrate their conceptual understanding, to relate smaller ideas to bigger ones, and to show that they grasp the overall significance of what they have learned. (pp. 10-11)

Let me depart from Conley's analysis for a moment to offer an analogy of his concepts. I mastered and taught a Korean martial art known as Kuk Sool Hapkido as a side interest. When teaching beginners, the master requires students to practice discreet or particular motions, called techniques, such as strikes and kicks, in preparation for their application. Two examples of application in the school setting are routines known as *forms*[28] (think of a choreographed dance) and sparring. The motions must be learned sufficiently (though not mastered at early stages of instruction – this comes with extensive practice) to combine them together to be applied to either the *forms* or to sparring. The *forms* teach the practitioner how to connect the techniques efficiently and harmoniously for the maximum use of force. Sparring – mock combat – teaches how to apply techniques to combat situations without exposing students to the lethality of real-world combat. In principle, this is what our educational system should be shooting for: Teach indiscreet pieces of knowledge in primary school so that they can be joined together and applied in a harmonious manner for learning-transfer at subsequent stages of education.

---

[28] Please see an example of a *form* displayed in two parts: Slow so the motions can readily be observed and then fast so application can be observed: https://www.youtube.com/watch?v=wvz0DF5TuYw This is the beginning basic form or "white belt" form.

Next Conley explains assessment methods that were common prior to formulating standardized tests – the one-size-fits-all approach – by the College Board in the early part of the 20th century. Schools used something akin to what we would currently call "performance assessment" ("referring to activities that allow students to show what they can do with what they've learned" p. 4) such as written essays and recitations, which was managed locally. It cannot be questioned that such assessment methods would develop language skills far above what we have today. If one is to be judged on how well one writes and speaks, one must practice writing and speaking extensively to develop these skills. The argument against traditional assessment methods was "These types of exams were not considered sufficiently scientific…." (p. 4)

To depart one more time from Conley: When testing students for competency in martial arts techniques, *forms*, and sparring, the primary method is demonstrable application – not multiple choice or true-false questions. There certainly are theories and abstract concepts (such as the use of "ki" or "chi" – life force) in more advanced martial arts but they are more a process of self-discovery, through proper guidance by the instructor, over an extended period of time.

Conley points out the flaws in the "scientific" assessment system:

> Unfortunately, the available testing technologies have never been complex or nuanced enough to make these types of predictions very successfully, and so assessments have been used (or misused, really) throughout much of the past century to categorize students…. Moreover, additional problems with such norm-referenced testing – designed to see how students stack up against one another – are readily apparent. In the first place, it is not clear how to interpret the results. By definition, some students will come out on top and others will rank at the bottom. But this is no reason to assume that the top scorers have mastered the given material (since they may just have scored a little less poorly than everybody else). Nor can it be assumed that the low scorers are in fact less capable (since, depending on where they happen to go to school, they may never have had a chance to study the given material at all). And, finally, even if they could be trusted to sort students into winners and losers, such tests would still fail to provide much actionable information as to what those students need to learn or do…. (pp. 4-5)

Next, Conley lists "Keys" that determine college readiness. 1) Key Cognitive Strategies; 2) Key Content Knowledge; 3) Key Learning Skills and Techniques; and 4) Key Transition Knowledge and Skills. These are things that students and teachers can act on to improve success.

Conley states, "[The Keys] model does not include certain factors, such as parental income and education level, that are strongly associated statistically with college success but which *are not actionable* by schools, teachers, or students. The point here is to highlight things that can be done to prepare students to succeed, not to list things that cannot be changed." (p. 9) This is so refreshing to hear! So many in the academic community wish to shift the responsibility onto other forces they have no control over – such as "social justice" issues with the attendant lip service they give it in order to placate the disenfranchised. This is the age-old blame-game strategy to divert attention away from real causes, which is the monopolization of education practices to serve the sector that controls it at the expense of all others.

Conley describes an innovative approach to assessment that needs serious analysis:

Assessments can be described as falling along a continuum, ranging from those that measure bits and pieces of student content knowledge to those that seek to capture student understanding in more integrated and holistic ways (as shown in Figure 2). But it is not necessary or even desirable to choose just one approach and reject the others. (p. 12)

Such approaches, in which a range of student assessment information is collected over time, permit educators to combine some or all of the elements on the continuum of assessments…. Doing so results in a fuller picture of student capabilities than is possible with any single form of assessment. And because this allows for the ongoing, detailed analysis of student work, it gives schools the option to assess their progress on relatively complex cognitive skills, which is very difficult to measure using occasional achievement tests. (p. 17)

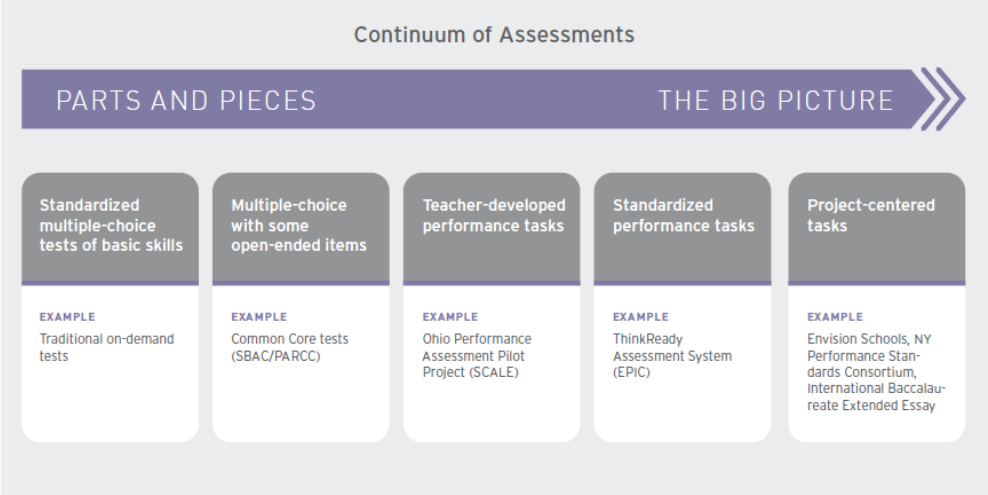


Figure 2

Conley addresses the various alternative assessment systems that developed such as: *multiple choice tests*, *common core tests*, *mastery learning* movement, *criterion-reference* testing, *outcomes-based education* or *academic standards* through statewide performance testing, *performance tasks*, *project-centered assessment*, *collections of evidence*, *metacognitive learning strategies assessments*, etc. While all of them may have some merit in them, individually, they fail students and society.

Conley provides a short analysis of each. It is recommended that Applied Education Foundation researchers, who may have need of this information, read his paper for greater insight into them. It can be accessed online via http://www.jff.org/sites/default/files/publications/materials/A-New-Era-for-Educational-Assessment-092414_0.pdf.

Conley's *Continuum of Assessments* shown above leads to his concept of a *System of Assessments*, which he defines as

> [A] comprehensive approach that draws from multiple sources in order to develop a holistic picture of student knowledge and skills in all of the areas that make a real difference for college, career, and life success. … Most important, it would avoid placing too much weight on any

> single source of data. In short, such a system would produce a nuanced and multilayered profile of student learners. … A system of assessments yields many more data points than does a single achievement test. Compared with the familiar connect-the-dots sketch of students' knowledge and skills, it offers a much more precise, high-definition picture of where they are, how far they've come, and how far they have to go in order to be ready for college and careers. (p. 20)

Conley concludes with a list of goals education policymakers will need to make. Here are several of the more significant goals:

- "Define college and career readiness comprehensively
- Take a hard look at the pros and cons of current state accountability systems
- Support the development of new assessments of deeper learning
- Learn from past efforts to build statewide performance assessment systems
- Take greater advantage of advances in information technology
- Build a strong base of support for a comprehensive system of assessments
- Determine the professional learning, curriculum, and resource needs of educators."

NWEA (2016) also addresses the need for multiple assessment methods, pointing out "Education stakeholders value assessments broadly, but views vary by assessment type and purpose." They share their findings on this point:

> ¾ of students and more than ½ of parents believe that students spend the right amount of time or too little time taking assessments. More than 7 in 10 teachers, principals and superintendents say students spend too much time taking assessments.
>
> Most teachers, principals and superintendents do not believe that state and federal policymakers understand the purpose of different types of assessment….
>
> While most parents understand that state accountability tests are used to evaluate school and district performance, many mistakenly believe that these types of tests are used to monitor student achievement and to inform instruction. More than 6 in 10 parents say their child's teachers rarely or never discuss their child's assessment results with them.

The Committee on the Foundations of Assessment provides a model to analyze how assessment is currently being used. In Part I, The Nature of Assessment and Reasoning from Evidence, page 37, the Committee considers the purposes of assessment by contrasting *achievement* from *aptitude* – "which has the purpose of predicting performance in some future situation." This report was not written to address *aptitude* but rather to consider *achievement*. They provide three subsections to address purposes of assessment with achievement in mind.

> **Assessment to Assist Learning**
> In the classroom context, effective teachers use various forms of assessment to inform day-to-day and month-to-month decisions about next steps for instruction, to give students feedback about their progress, and to motivate students. …
>
> **Assessment of Individual Achievement**

> Another type of assessment used to make decisions about individuals is that conducted to help determine whether a student has attained a certain level of competency after completing a particular phase of education, whether it be a classroom unit or 12 years of schooling. …
>
> **Assessment to Evaluate Programs**
> Another common purpose of assessment is to help policy makers formulate judgments about the quality and effectiveness of educational programs and institutions…. Assessments are used increasingly to make high-stakes decisions not only about individuals, but also about institutions.

These three assessment summaries require us to rethink what is to be taught, how it is to be taught, and how a diverse offering will be assessed.

Rutter (2015) points out that postsecondary educators tend not to assess students in appropriate ways, but, rather, simply rank

> students against a rubric or their peers on the basis of a quiz, a problem set, a multiple choice or essay exam, or a report or research paper. But as any K-12 educator could explain, assessment can serve many other, even more important, purposes, each of which demands very different kinds of assessment exercises. What are the functions that assessment can serve? Here are nine:
>
> - To establish a reference point
> - To diagnose a student's needs and strengths
> - To individualize instructional pathways
> - To motivate student learning
> - To prompt metacognition
> - To measure progress
> - To provide feedback and guidance
> - To establish accountability
> - To evaluate a student's performance against a rubric or against other students.
>
> …Typically, assessments correspond to a particular dimension of Bloom's Taxonomy, the classification of thinking skills developed by educational psychologist Benjamin Bloom in 1956, which distinguishes between recall, comprehension, application, analysis, synthesis, and evaluation.

While these are good reference points for assessment purposes, they cannot be looked at as serving only academic purposes. "To diagnose a student's needs…" must include discovering the individual's talents and to help guide the student in that direction, which Rutter touches on when he states, "Assessments can be traditional or authentic – that is, those that are academic and those that involve the skills and knowledge needed to perform a 'real world' task." Rutter expands upon this where he provides:

> Performance-based assessment offers a valuable alternative (or supplement) to the standard forms of student evaluation. Performance-based assessment requires students to solve a real-world problem or to create, perform, or produce something with real-world application. It allows an instructor to assess how well students are able to use essential skills and knowledge, think

> critically and analytically, or develop a project. It also offers a measure of the depth and breadth of a student's proficiencies.
>
> Performance-based assessment can, in certain instances, simply be an example of what Bloom's Taxonomy calls application. Thus, a student or a team might be asked to apply knowledge and skills to a particular task or problem. But performance-based assessment can move beyond Bloom's Taxonomy when students are engaged in a project that requires them to display creativity and that results in an outcome, project, or performance that is genuinely new. The more sophisticated performance assessments involve research, planning, design, development, implementation, presentation, and, in the case of team-based projects, collaboration.
>
> … If an assessment is to be meaningful and fair, it must be "valid" and "reliable" – that is, the results must be accurate and consistent. … Many assessments are anything but valid and reliable. … If performance-based assessments are to be fair, valid, and reliable, it is essential that there is an explicit rubric that lays out the criteria for evaluation in advance.

National Research Council (2000) offers insight into the role assessment plays in education:

1. "The roles for assessment must be expanded beyond the traditional concept of testing. … Given the goal of learning with understanding, assessments must tap understanding rather than merely the ability to repeat facts or perform isolated skills." (p. 19)
2. "Assessment for purposes of accountability (e.g., statewide assessments) must test deep understanding rather than surface knowledge. Assessment tools are often the standard by which teachers are held accountable. A teacher is put in a bind if she is asked to teach for deep conceptual understanding, but in doing so produces students who perform more poorly on standardized tests. Unless new assessment tools are aligned with new approaches to teaching, the latter are unlikely to muster support among the schools and their constituent parents. This goal is as important as it is difficult to achieve. The format of standardized tests can encourage measurement of factual knowledge rather than conceptual understanding, but it also facilitates objective scoring. Measuring depth of understanding can pose challenges for objectivity. Much work needs to be done to minimize the trade-off between assessing depth and assessing objectively." (p. 20)
3. "Tests often reinforce memorizing rather than understanding." (p. 24)
4. "Assumptions about transfer accompany the belief that it is better to broadly 'educate' people than simply "train" them to perform particular tasks." Measures of transfer play an important role in assessing the quality of people's learning experiences. Different kinds of learning experiences can look equivalent when tests or learning focus solely on remembering (e.g., on the ability to repeat previously taught facts or procedures), but they can look quite different when tests of transfer are used. Some kinds of learning experiences result in effective memory but poor transfer; others produce effective memory plus positive transfer. (p. 51)

As a postscript, Western Governors University (Meyers, 2019), being an applied education-oriented school, provides instruction, and perhaps a template, in what goes into organizing and structuring a model assessment system. Meyers states:

As a competency-based university, assessment is the backbone of everything we do. Quality, **industry-verified**[29] tests allow us to confidently classify an individual as competent and prepared for workforce entry. (Emphasis added.) Western Governors University (WGU) is different from the rest of academia in that assessments are central to how we validate skills and competencies. The traditional credit hour means little at WGU. Instead, students are able to demonstrate mastery through the use of high-quality assessments – objective tests, papers, simulations, clinicals or presentations. … What follows is a glimpse into how WGU's rigorous process for objective assessments ensures value to students and the employers who hire WGU graduates. …

Linking Skills to Professions

[W]e perform what is broadly referred to as a Job Task analysis. This is how we ensure that the skills we measure matter on-the-job. Typically, surveys are conducted asking professionals to rate various skills in terms of how important they are to, and how frequently they are used on, the job. We have created a unique process whereby all underlying skills are connected across all courses and degree programs….

Specifying What's on the Test

Next, test specifications are created. These specifications comprise a test description … and a test blueprint. …

The test blueprint is then made available to students and instructors so that they have solid expectations about how they will be evaluated. Students should not have to guess what is considered important when preparing for an exam. We are transparent so that students know what is important and how to prepare for their assessment, thereby reducing test anxiety. …

Writing Items

After the blueprint is created, items are written. While most people, particularly teachers and professors, think they can write good items, it isn't as easy as it sounds and WGU holds high standards in this regard. There is some science behind how to do it appropriately. There are entire books written and courses taught on this subject – just because one is a good teacher does not mean one can write good items. Poor items might have no correct answer, multiple correct answers, be vague and confusing, or be so specific they lack relevance. Only a subset of written items (often around 50%) pass all the tests we put them through and meet all the requirements for use. Assessment developers work with subject matter experts to construct a set of items to meet the test purpose. …

Review, Review and Review

Items then go through a number of quality reviews. Subject matter experts review the items across multiple rounds to make sure the content is accurate, relevant and in concert with curricular materials. Assessment experts review the items for clarity, and fairness and sensitivity. Copyeditors review the items for grammar punctuation, spelling and style.

[29] Industry associations are key to formulating education for the needs of individuals and society. Education institutions should provide the delivery system and support in formulating curricula but they should not decide content or context.

Test the Test

After an iterative process[30] of review and revision, questions need to be tested with students or a similar group. This is called field testing and the purpose is to further evaluate the quality of the items. The best way to collect this data is to fold these items in with real items on some test without the participants knowing which ones count and which do not.

Human Attributes That Influence Success in Life

Conley (2014) provides a look at what it takes to succeed in life; whether that life is immersed in the educational system, economic activity, public arena, etc. He points to qualities of character that make the real difference.

> [P]ublic interest has surged, of late, in the role that perseverance, determination, tenacity, and grit can play in learning. So, too, has the notion of academic mindset struck a chord with many practitioners who see evidence daily that students who believe that effort matters more than innate aptitude are able to perform better in a subject. And researchers are now pursuing numerous studies of students' use of study skills, their time management strategies, and their goal-setting capabilities.
>
> In large part, what makes all of these metacognitive skills so appealing is the recognition that such things can be taught and learned, and that the evidence suggests that all are important for success in and beyond school. (p. 18)

Conley also included necessary attitudes, habits, behaviors, and beliefs to the list of necessary attributes for success. (p. 19) Put this all together and we can observe some of the ingredients that help contribute to and define ethics – human attributes post-modernists marginalize, unless it is their brand of "ethics," which is always a moving target.

Cohen et al. (2017) provide a deeper understanding of individual motivational forces.

> In a nationwide survey of high school dropouts, 69% said that school had not motivated or inspired them to work hard. In fact, many of the students who remain in school are not motivated or inspired either, and the more time students spend in K-12 education the worse it gets. This lack of motivation to do well in school represents a serious loss of human potential, with implications for students' well-being later in life and for our country's future economic growth. What prevents students from working hard in school? Is it something about them or is it something about school?
>
> …Most educational reforms focus on curriculum and pedagogy – *what* material is taught and *how* it is taught. However, curriculum and pedagogy have often been narrowly defined as the academic content and students' intellectual processing of that material. Research shows that this is insufficient. In our pursuit of educational reform, something essential has been missing: the

[30] An iterative process is a process for calculating a desired result by means of a repeated cycle of operations. An iterative process should be convergent, i.e., it should come closer to the desired result as the number of iterations increases. http://pespmc1.vub.ac.be/ASC/ITERAT_PROCE.html

> psychology of the student. **Psychological factors – often called *motivational* or *non-cognitive* factors – can matter even more than cognitive factors for students' academic performance**. … Educators, psychologists, and even economists recognize the importance of non-cognitive factors in achievement both in school and in the labor market. … The research … shows that educational interventions and initiatives that target these psychological factors can transform students' experience and achievement in school, improving core academic outcomes such as GPA and test scores months and even years later. (Emphasis added)
>
> … Past attempts to motivate students by promoting positive beliefs have included the self-esteem movement of the 1990s, which tried to motivate students by making them feel good about themselves, their abilities, and their prospects of success in school. Unfortunately, the self-esteem movement had the erroneous view that telling students they were smart or talented would raise their self-esteem and motivate them to do well in school. In fact, research has now shown that well-intended practices, such as praising students' intelligence or talent (as opposed to their efforts or strategies), often backfire. (p. 2)

Here is evidence showing that assessment tests may measure some academic ability but motivation is frequently more important. I have known numerous individuals who were straight "A" students, but did not amount to much in life due to a lack of motivation. One may think that it takes great effort to attain a GPA in the upper 3s or a 4.0, but this is not the case for one who has excellent memory retention and recall abilities. Such minds are more like a computer than a reasoning being. Therefore, school was easy and since school is promoted as "the measure of all things," opportunities should be offered to such individuals on a silver platter – shouldn't they? Unfortunately, the educational establishment misinformed them; this is not how the real world works.

If students possess the motivational attributes described by Cohen et al., content knowledge drops in importance since individuals with such attributes can always acquire subject matter content on their own if they have mastered the ability to do self-directed learning and research (Schwartz et al., 2005, refer to it as "preparation for future learning" or PFL). However, this in no way implies that content knowledge is irrelevant or of no use – it's simply a matter of setting priorities. Therefore, it becomes the teacher's priority to help students acquire the attributes Conley mentions – along with the ability to do self-directed research – so they can become self-learners with less dependence on institutional environments since the nature of institutions is to erect barriers in order to protect factional interests. Prior to the 20th century bureaucratically controlled educational State systems, self-learning and self-discovery was understood to be the goals of education. The idea was not to lead individuals to a dependence on institutions as we have it now, but rather, to independence and self-determination.

AACC (2015) names some highly successful individuals who did not fit within the academic environment. Of course, there is an extensive list of such people. Andrew Carnegie (1902, pp. 109-14) points out that too much education conditions people for academia but not for the world of work nor for positions of leadership. Therefore, highly educated people were never able to become captains of industry. So, while higher education is the answer for some callings, it is not for all, nor even the majority. Therefore, we must provide alternatives for all walks of life. To do otherwise demonstrates prejudice toward those people and those walks of life academia has not anointed as worthy.

As an aside, Bucholz (2003, lecture 27) informs us that William Shakespeare possessed only a grammar school education, yet many consider him the most eloquent and creative master of the English language to date. This demonstrates that innate talent transcends formal education. That is, education may contribute to the development of one's talents, but it cannot replace it. Furthermore, one's talents can do quite nicely without a formal education as the endless array of historic examples demonstrate.

AACC summarizes the findings of research done by Gaertner and McClarty of Pearson's Research & Innovation Network.

> Gaertner and McClarty analyzed longitudinal data from more than 11,000 students in the National Education Longitudinal Study (NELS) of 1988. They used 140 middle school variables to create six factors: achievement, behavior, motivation, social engagement, family circumstances and school characteristics. They then used these middle school factors to predict college readiness and college outcomes.
>
> They found some interesting results:
>
> - With respect to college readiness, all six factors were valuable predictors. As we might expect, academic achievement was the strongest predictor (accounting for 17 percent of the variation in college readiness), but motivation (15 percent) and behavior (14 percent) followed closely. Together, motivation and behavior contribute more to college readiness than achievement on its own.
>
> - With respect to college success, these six middle school factors predict college grades (cumulative GPA) and graduation better than the ACT or the SAT. In fact, using the six factors nearly doubles our ability to predict college graduation, compared with using test scores alone. Most notably, these predictions arrive in eighth grade — three years before the SAT or the ACT.

The non-cognitive human attributes that determine success in all aspects of life demands our in-depth attention as it relates to the bridge between illiteracy and literacy and the subsequent bridge that leads to social and economic success. I believe Goleman (1995) offers one of various human dynamics and attributes that require in-depth analysis – emotional stability – that can lead us toward success. He states,

> The last decade … has … seen an unparalleled burst of scientific studies of emotion. Most dramatic are the glimpses of the brain at work, made possible by innovative methods such as new brain-imaging technologies. They have made visible for the first time in human history what has always been a source of deep mystery: exactly how this intricate mass of cells operates while we think and feel, imagine and dream. This flood of neurobiological data lets us understand more clearly than ever how the brain's centers for emotion move us to rage or to tears, and how more ancient parts of the brain, which stir us to make war as well as love, are channeled for better or worse. This unprecedented clarity on the workings of emotions and their failings brings into focus some fresh remedies for our collective emotional crisis. … Now science is finally able to speak with authority to these urgent and perplexing questions of the psyche at its most irrational, to map with some precision the human heart.

> This mapping offers a challenge to those who subscribe to a narrow view of intelligence, arguing that IQ is a genetic given that cannot be changed by life experience, and that our destiny in life is largely fixed by these aptitudes. That argument ignores the more challenging question: What *can* we change that will help our children fare better in life? What factors are at play, for example, when people of high IQ flounder and those of modest IQ do surprisingly well? I would argue that the difference quite often lies in the abilities called here *emotional intelligence*, which include self-control, zeal and persistence, and the ability to motivate oneself. And these skills, as we shall see, can be taught to children, giving them a better chance to use whatever intellectual potential the genetic lottery may have given them. (pp. xi-xii)

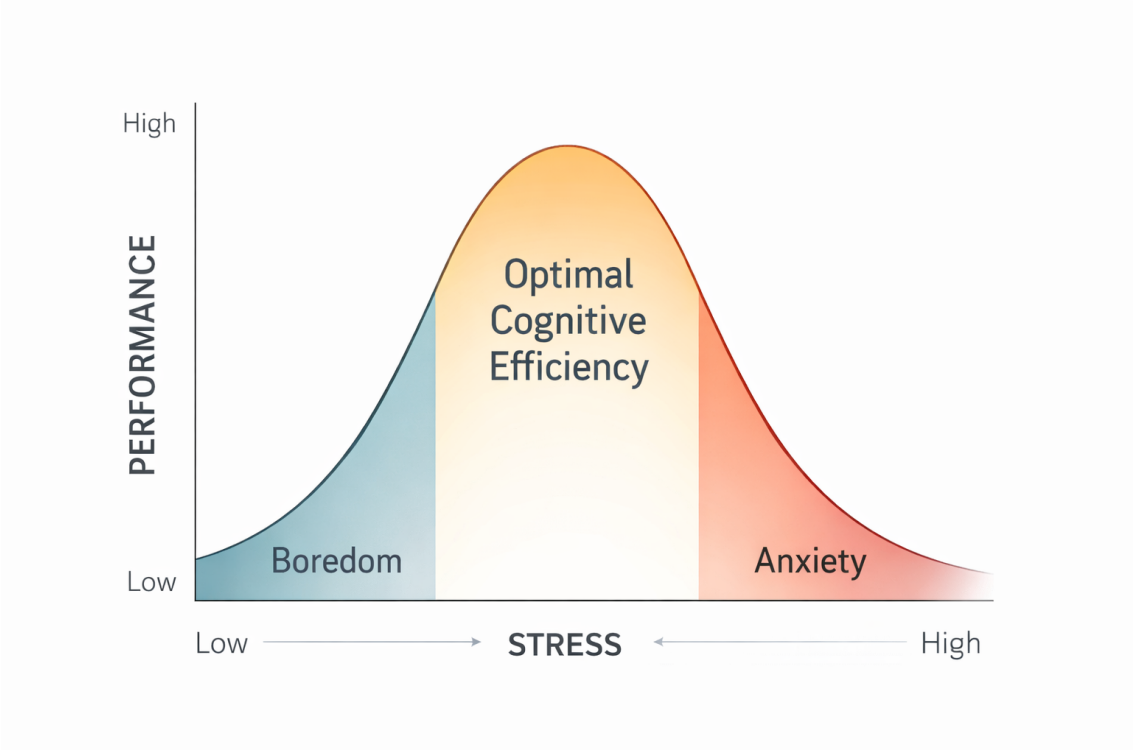
In 2006, Goleman published another work, but this time on the subject of *social intelligence*, which is another human attribute that requires further investigation for educational purposes. This work is meant to broaden our understanding of intelligence that was begun in *Emotional Intelligence* by analyzing human interaction and the effects it has on individuals. Let's look at what he offers in Chapter 19: The Sweet Spot for Achievement. He cites an article, "The Biology of Being Frazzled" from the academic journal *Science*:

> "Frazzle" is a neural state in which emotional upsurges hamper the workings of the executive center. While we are frazzled, we cannot concentrate or think clearly. That neural truth has direct implications for achieving the optimal emotional atmosphere both in the classroom and the office.
>
> From the vantage point of the brain, doing well in school and at work involves one and the same state, the brain's sweet spot for performance. The biology of anxiety casts us out of that zone for excellence.
>
> "Banish fear" was a slogan of the late quality-control guru W. Edwards Deming. He saw that fear froze a workplace: workers were reluctant to speak up, to share new ideas, or to coordinate well, let alone to improve the quality of their output. The same slogan applies to the classroom – fear frazzles the mind, disrupting learning.
>
> The basic neurobiology of frazzle reflects the body's default plan for emergency. When we are under stress, [hormones roar] into action, preparing the body for crisis. Among other biological maneuvers, the amygdala[31] commandeers the prefrontal cortex, the brain's executive center. This shift in control to the low road favors automatic habits, as the amygdala draws on knee-jerk responses to save us. The thinking brain gets sidelined for the duration; the high road moves too slowly.
>
> As our brain hands decision-making over to the low road, we lose our ability to think at our best. The more intense the pressure, the more our performance and thinking will suffer. The ascendant amygdala handicaps our abilities for learning, for holding information in working memory, for reacting flexibly and creatively, for focusing attention at will, and for planning and organizing effectively. We plunge into what neuroscientists call "cognitive dysfunction."

[31] The "area in the midbrain that triggers the fight, flight, or freeze response to danger." (p. 14)

> … The greater the anxiety we feel, the more impaired is the brain's cognitive efficiency. In this zone of mental misery, distracting thoughts hijack our attention and squeeze our cognitive resources. Because high anxiety shrinks the space available to our attention, it undermines our very capacity to take in new information, let alone generate fresh ideas. Near-panic is the enemy of learning and creativity.
>
> … [A]s we become preoccupied by, say, worry or resentment, our mental agility sputters. Likewise, when we are sad, activity levels in the prefrontal cortex drop and we generate fewer thoughts. Extremes of anxiety and anger on the one hand, and sadness on the other, push brain activity beyond its zones for effectiveness.
>
> Boredom fogs the brain with its own brand of inefficiency. As minds wander, they lose focus; motivation vanishes. In any meeting that has gone on too long (as most do), the vacant eyes of those trapped at the table will betray this inner absence. And we all remember days of [boredom] as students, absently staring out the window. (pp. 267-68)

"Frazzle" is something most students feel when taking tests – particularly high stakes tests. Therefore, we have a "comedy of errors" from two fronts: 1) The current testing methodology does not predict success in school nor in the real world, and 2) The pressure students experience regarding the expectations to perform on these tests, if they are to hope for success in life, frazzle them so they are unable to perform at their optimum. There is most certainly a contradiction from these two fronts that establishes a whirlwind of confusion. So, between the uselessness of these tests and the stress they engender, we set our youth up for emotional instability at a very young age. Many individuals never recover from the educational culture's damage wrought during the stages of development.



Goleman (2006) offers imagery that sheds light on the "frazzle" effect (see above):

> Plotting the relationship between mental adeptness (and performance generally) and the spectrum of moods, creates what looks like an upside-down U with its legs spread out a bit. Joy, cognitive efficiency, and outstanding performance occur at the peak of the inverted U. Along the downside of one leg lies boredom, along the other anxiety. The more apathy or angst we feel, the worse we do, whether on a term paper or an office memo.
>
> We are lifted out of the daze of boredom as a challenge piques our interest, our motivation increases, and attention focuses. The height of cognitive performance occurs where motivation and focus peak, at the intersection of a task's difficulty and our ability to match its demand. At a tipping point just past this peak of cognitive efficiency, challenges begin to exceed ability, and so the downside of the inverted U begins.
>
> We taste panic as we realize, say, we've procrastinated disastrously long on that paper or memo. From there our increasing anxiety erodes our cognitive efficiency. As tasks multiply in difficulty and challenge melts into overwhelm, the low road becomes increasingly active. The high road frazzles as the challenges engulf our abilities, and the brain hands the reins to the low road. This neural shift of control from the high to the low road accounts for the shape of the upside-down U. (pp. 270-71)

What Goleman's two works provide, reveals a gulf the size of the Grand Canyon between what educators understand of human dynamics & cognition and the results a free society demands of a public education system. And it all starts with the horse pushing the cart; that is, the test-making and accreditation industries dictate what schools are to teach regardless of individual needs, societal needs, and the end results society expects. Public school teachers' hands are tied and are forced to work within the confines of a dysfunctional system, which causes some teachers to behave in dysfunctional ways since they are required to do things that are contrary to common sense, compassion, decency, and equity. My heart goes out to them! I hope that relief will be forthcoming shortly.

Challenging the Conventions of Grading and Reporting

Guskey (2015) analyzes what he calls "Essential Beginning Steps" for reform:

> Efforts to make grading and reporting more effective and to bring it more in line with other reform initiatives can take many different directions. The efforts that succeed, however, typically begin with three initial steps.
>
> 1. Clarify the purpose of grading and reporting.
> 2. Align all classroom and school policies and practices with the stated purpose.
> 3. Ensure proposed changes are supported by strong research evidence. (pp. 4-5)
>
> In any educational setting where the central purpose is to encourage student learning, grading and reporting should always be done in reference to specific learning criteria. Because normative[32] criteria and grading on the curve tell nothing about what students have learned or are

[32] "[The] process of assigning grades based on a student's relative standing among classmates is referred to as *normative-based grading*, or more familiarly as *grading on the curve*." (Guskey, 2015, p. 47)

> able to do, they provide an inadequate description of student performance. Plus, they promote unhealthy competition, destroy perseverance and other motivational traits, and are generally unfair to students.
>
> … Until teachers, professors, and instructors at all levels precisely identify what students are expected to learn, articulate the criteria by which student learning will be judged, and clearly communicate these criteria to students, grading will remain an arbitrary and highly subjective process that diminishes learning and victimizes more students than it helps. (p. 58)

Guskey then raises an incredibly revealing question that he poses to teachers in deciding their purpose in the educational community:

> Those who choose education as their profession must answer one basic question about their purpose. How they answer this question will largely determine the direction of their career, how they go about their work, and how they judge their success. No other question about their role as an educator will be more important.
>
> That one basic question is this: Is my purpose to *select* talent, or is it to *develop* talent? The answer must be one or the other, because there is no in-between.
>
> If you decide your purpose as an educator is to *select* talent, then you must work to accentuate and maximize the differences among students. In other words, on any measure of student learning, you must try to achieve the greatest possible variation in students' scores. (p. 59)

This means whether all students in a class do extremely well in comprehending material and demonstrating competency, some must be A students and some must receive F's since a bell curve demands this of such a teacher, even if the differences are minute. This also means that if all of one's students are utterly incompetent at what's been taught, there will still be A students and F students; again, even if the differences are minute. This is an evil of the highest order since it is a direct attack on most citizens so that a "select" few may rise above the rest, even though there is no correlation – and perhaps an inverse correlation – between academic performance and real-world leadership abilities. It is an attempt to create a new world order of academic aristocrats who will control the reins of empire even though they have no idea what they are doing, as Woodrow Wilson and Barak Obama demonstrated.

Taking the principle of *selectivity* to the next level, Guskey points out that

> College entrance examinations like the ACT and SAT … are designed specifically for the purpose of selection. They are developed to accentuate the differences among students in order to facilitate decisions about admission to selective colleges…. As such, they are not created to measure what students have learned and are able to do. They also do not help develop students' talents or abilities or offer evidence on how well students were taught. In fact, scores on assessments like the ACT and SAT tend to be influenced more by students' socioeconomic status (SES) than by the quality of instruction students receive. (p. 60)

Guskey continues:

> The best means known for maximizing differences in student learning is poor teaching. Nothing does it better. If you want to accentuate the differences among students, then teach them as poorly as possible. A few students will be able to direct their own learning and will achieve at a high level, regardless of what the teacher does.[33] The vast majority of students, however, need guidance and direction in their learning. To learn well, they need to engage in structured learning opportunities and receive support from their teachers. Without such opportunities and support, they are likely to learn very little. The differences in student learning will thus be maximized, and this variation will be evident in any measure of learning.
>
> … The fundamental issue in computing students' class rank is: Why do we do it? Why do we believe it is necessary to rank-order all students in every graduating class? … What purpose does the ranking of graduating students serve?
>
> If we go back to the original question about our purpose as educators – specifically, is our purpose to *select* talent or to *develop* talent – then the answer is clear. Rank-ordering the students in every graduating class has nothing to do with developing students' talent. Rather, it is unquestionably about selecting talent.
>
> The process of determining class rank does not help students achieve more or reach higher levels of proficiency.[34] With the possible exception of the top-ranked student, it also does nothing to enhance students' sense of self-worth, their confidence as learners, or their motivation for learning. In fact, evidence indicates that it actually may accomplish quite the opposite and diminish students' motivation. If we say our purpose is truly to develop the talents of students, then the process of computing class rank is unmistakably counter to that purpose. (pp. 59-62)

To emphasize the point that the education establishment's assessment paradigm does not really select for talent, Guskey reflects on the work done by Karen Arnold (1995), *Lives of Promise: What Becomes of High School Valedictorians*.

> … Arnold found that while most of the valedictorians were successful, well-adjusted, and psychologically healthy, they were seldom at the head of the class in their careers. Most chose conventional careers as accountants, physicians, lawyers, engineers, physical therapists, and health care professionals, and worked well within the system. But few were risk takers or mold breakers. They also were unlikely ever to change the system. They worked hard and followed the rules but rarely proposed innovations or explored unfamiliar areas. Arnold summarized the results, saying, "Just because they could get A's doesn't mean they can translate academic achievement into career achievement."
>
> So the question this leaves for educators is this: Do current policies for selecting the class valedictorian foster development of the traits we most value in students? And if not, what policies might? (p. 67)

---

[33] We see this in all endeavors. Those who possess talents in the various "intelligences" – better to say talents and abilities – don't necessarily require formal instruction. Knowledge can be gained from many sources and public schools are not necessarily the best place to get it.

[34] Hence the reason employers lament a poorly educated workforce.

In Guskey's chapter *Final Thoughts*, he concludes:

> What should be apparent is that grading is a difficult and troublesome area. It's also an area in which most current policies and practices are bound more by tradition than by evidence of effectiveness. We don't do what we do in grading because we have thought about it in depth or considered thoroughly the intended as well as unintended consequences of our practices. More often than not, we do it simply because "we've always done it that way."
>
> There is much about grading and reporting we still don't know. We don't yet have sufficiently conclusive evidence to identify precisely what is truly *best* practice. It remains an area of education ripe for carful study and thoughtful investigation. (p. 109)

Lubischer (2023) reinforces Guskey's analysis. On the evils of "academic rigor," she writes:

> We all know the story of rigor: "Rigor is how we prepare students for the real world, and our job as faculty is to bring rigor to our classes." In this story, rigor is the hero and faculty members are its enforcers.
>
> The story of rigor is an example of a stock story, a narrative told over and over again that teaches and reinforces one view of reality. Stock stories provide a simple view of the world, told from a singular, dominant perspective, one that obscures other, messier and more complex realities.
>
> Rigor, according to this stock story, is core to the very existence of higher education, because it enables us to determine who is worthy. We have all heard some version of this sentiment: "To be of value, a course must be rigorous." In other words, a course is of value only if it is hard enough to enable us to rank-order the students. Higher education serves as a gateway to opportunity, and rigor is how we determine who can pass through the gate.
>
> And students are not the only ones being subjected to rigor-informed judgment – faculty members are as well. "She gets good student evaluations only because her courses aren't rigorous." In my experience, this critique is more often aimed at faculty not on the tenure track or assistant professors who some colleagues think do not belong.[35]
>
> As a neurobiologist and an educator guided by a science of learning framework, I do not dispute that rigor can play a role in learning. The role of rigor, however, is misrepresented by this stock story – so much so that Jordynn Jack and Viji Sathy make a strong case for no longer using the word "rigor" in conversations about higher education. The misrepresentation of rigor becomes clear if we move past the dominant stock story of rigor to acknowledge and examine the messier reality it conceals.

[35] Most assuredly, an elitist view in attempt to elevate educators as the contemporary aristocracy. This is a very dangerous game since it leads to something akin to serfdom. After all, they are attempting to create winners and losers with wealth and power going to the winners.

Learning (not rigor) is what prepares students for the real world, and our job as faculty members is to strategically challenge our students and help them engage with that challenge to enable their learning. In this concealed story, learning is the hero of higher education and faculty members are its facilitators, while rigor for the sake of rigor is the antihero.

Learning happens when the learner directs their neural attention system toward a cognitive challenge and engages with that challenge in a manner that results in consolidation of the new learning. This creates representative neural circuits that can be further strengthened through continued use of what was learned. To the extent that we align cognitive challenges with course learning outcomes and assessments, all students have an opportunity to learn and to demonstrate their learning.

Not all challenges, however, serve to promote learning. A course or curriculum can be rigorous in ways that have little to do with learning (some of which are detailed by Kevin Gannon): give assessments that do not align with what was actually taught in the course, test students on a small sample of a massive amount of material, provide no feedback along the way to a single high-stakes assessment, enforce inflexible deadlines and time constraints, include requirements for no good reason, and be sure to let students know that they do not really belong in the course or in the major ("If you don't do well in this course, you probably shouldn't be in science").[36]

Anything that makes a course harder for at least some students makes that course more rigorous, whether or not those challenges actually support learning. This sort of rigor is much like riddles told by storied gatekeepers such as the Sphinx of Thebes— arbitrary barriers that support the myth of meritocracy by giving the gatekeeper an irrelevant reason to let only a chosen few pass. Rigor in the form of arbitrary barriers does not support learning, and arbitrary barriers are detrimental to student well-being.

The stock story of rigor holds that a course is rigorous if enough students fail or withdraw, regardless of the reasons. When not enough students fail, we hear complaints from the most vigorous defenders of rigor about grade inflation and comments like this: "If no students failed, then it obviously wasn't a rigorous class." It is critical to the gateway function of higher education that some students fail and that their failure is documented on the transcript.

Using student failure as a measure of rigor assumes that some win only if others lose, thereby reinforcing the value of competition. But the amount of learning in a class or curriculum is unlimited, so why should it be competitive? Earning a grade is competitive only if faculty are meant to be gatekeepers (defenders of rigor) rather than educators (facilitators of learning). Faculty should be allowed to be educators, to carefully design appropriately challenging courses and practice pedagogies of care, enabling all of our students to learn.

[36] And yet the academic community does not accept much of the blame that rests at their feet for the mental health crisis plaguing high school and college students.

Another messy reality is that the gradations inherent in grades (it's right there in the name) do not reliably indicate who learned more and who learned less, nor do grades indicate which students will retain their learning beyond the end of finals week. A premed student once told me that they missed a course grade by one point out of 1,000 in a class—a 0.1 percent determination of how well they had learned compared to other students. We do not have that degree of accuracy in measuring learning, yet final course grades and transcripts give the power to pretend as much.

A simple letter grade on a transcript cannot tell us why the grade is what it is or the story of how it came to be. What role did life events play? Was the student working long hours, struggling through stereotype threat, learning and communicating in a nonnative language, dealing with relationship drama, caring for a sick parent, or facing their own health issues? How did life impact their grade? External influences on individual student academic performance cannot be measured reliably, so we ignore them—or perhaps try to mitigate students' challenges as best we can (e.g., through flexible deadlines). Then we post a simple grade as a far-too-oversimplified summary that cannot accurately reveal what a student experienced and learned in our course.

The letter grades we post then become part of a transcript that a student may be required to share with a desired graduate or professional program or employer. The transcripts that higher education creates not only list accomplishments but also document failures in a permanent manner: our transcripts are academic rap sheets.

We can do better.

We can reject rigor for the sake of rigor as nothing more than traditional gatekeeping: stop rewarding faculty members for high course failure rates and stop stigmatizing faculty for high learning rates.
We can write a transformative story for all of higher education by committing individually and institutionally to 1) course and curricular design that incorporates challenges appropriate to learning, 2) courses and classrooms transformed through pedagogies of care (e.g., inclusive teaching, trauma-informed pedagogy), and 3) the elimination of policies and practices that create arbitrary barriers to student success.

Finally, we can turn transcripts into records of achievement rather than rap sheets. In a brighter, more equitable future, I envision a transcript that lists courses completed successfully (no letter grades) and degrees earned. This reimagined transcript—let's call it the achievement report— will organize student accomplishments by area of study rather than chronologically, because the topic matters more than the term of completion. The achievement report will indicate not just course titles, but also which courses emphasized specific skill sets that employers highlight as being desirable, such as communication, problem-solving, teamwork and creativity.
Replacing the rap sheet version of a transcript with an achievement report that emphasizes students' academic accomplishments will require leadership in institutions of higher education across the U.S. But if we truly want to support student learning and well-being, we will find a way to demonstrate what a student has learned and not where

> they stumbled. In doing so, we will help our students succeed both during their college journey and after they graduate.[37]

This article points to the fact that students aren't learning much in college. Rather, they are treading water trying to keep afloat in a sea of arbitrariness, confusion, and subservience to "the masters." One should, therefore, not be surprised by employers' vote of no-confidence in the abilities of graduates.

Janio's article, "Completion is Not Competence: Why Accountability Requires a Definition of Learning" (2026), explains how the educational establishment has used proxies for the concept of competence. "Learning … is … inferred from grades, credit accumulation or aggregated reports whose criteria vary across programs and disciplines. We count what moves through the system. We struggle to specify what changes because of it. … [These measures are not] designed to define and verify what learning must consist of in practice."

Military Assessment Perspective

The U.S. Army (*Army Talent Assessment Strategy*, Dept. of the Army, Jan. 31, 2024.) touches briefly on assessment practices:

> Assessment practice and test preparation can improve or change test scores in critical ways. It is well known that some individuals are "better" at taking tests and assessments than others. Test preparation courses (e.g., for the Graduate Record Examination (GRE), or the Future Soldier Preparatory Course for ASVAB) are commonly accepted practices and take advantage of this knowledge to teach people how to improve their scores. Assessment scores may change (and usually improve) for these reasons when an individual takes the same or similar assessments multiple times. These changes are commonly referred to as 'practice effects.' Other reasons are related to more detailed knowledge about test-taking strategies, guessing strategies, and understanding how scoring of the assessment is done. These are all referred to as test-taking skills. Most test preparation courses use a combination of practice effects and test-taking skills to improve individuals' test scores. Test preparation courses can inadvertently introduce assessment bias or unfairness into the assessment if test preparation is available for some subgroups but not others. The Army should consider the extent to which it makes available test preparation courses to individuals to remove this as a significant influence on assessment scores.

This indicates that what the test is assessing isn't actual reasoning abilities; but rather, it's a measurement of data memory and recall. While data recall has its place, it doesn't measure reasoning abilities.

The Army (*Adult Learning in the Military Context (2024)*, Chapter 7, "Learning Assessments and Evaluation" pages 262-93, offers insight into considering what to assess, the purpose of assessments, and how to evaluate programs.

We begin with the section *Definitions of Tests, Assessments, and Evaluation*.

---

[37] Jane Lubischer is associate department head in the Department of Biological Sciences at North Carolina State University. She wrote this article with support from the Howard Hughes Medical Institute.

The committee uses a number of definitions set out in the *Standards for Educational and Psychological Testing (The Standards)* (AERA et al., 2014), a widely recognized document providing "a basis for evaluating the quality of [testing] practices" (p. 1). These definitions include the following:

- **Assessments**: "Any systematic method of obtaining information, used to draw inferences about characteristics of people, objects, or programs; a systematic process to measure or evaluate the characteristics or performance of individuals, programs, or other entities, for purposes of drawing inferences; sometimes used synonymously with *test*."
- **Tests**: "An evaluative device or procedure in which a systematic sample of test taker's behavior in a specified domain is obtained and scored using a standardized process."
- **Program evaluation**: "The set of procedures used to make judgments about a program's design, its implementation, and its outcomes"; policy studies and accountability are broader, but related concepts. Program evaluation uses the data collected during the assessment process.
- **Validity**: "The degree to which evidence and theory support interpretations of test scores for proposed uses of tests" (and note that "it is the interpretations of test scores for proposed uses that are evaluated, not the test itself").

The committee notes that the Educational Testing Service's (ETS) (2014) *Standards for Quality and Fairness* defines the terms "test" and "assessment" as synonymous and defines test as "a systematic sample of behavior taken to allow inferences about an individual's knowledge, skill, ability, or other attribute." Assessment and evaluation of adult learning can serve a variety of purposes. Purpose, therefore, is a critical issue in determining which assessment approaches are most effective, as different approaches are effective for different purposes.

High- and Low-Stakes Assessment

One important dimension of test use concerns the stakes associated with a test's scores. High-stakes tests are those "used to provide results that have important, direct consequences for individuals, programs, or institutions involved in the testing." One high-stakes assessment that many enlisted personnel encounter early in their career is the Armed Services Vocational Aptitude Battery (ASVAB). This assessment is high-stakes because it directly informs their classification into their military occupational specialty. Other examples include promotion tests, recognition and award tests, and end-of-course summative pass/fail assessment. For these situations, test scores have important, direct consequences for the individual test taker.

Low-stakes tests are those "used to provide results that have only minor or indirect consequences for individual[s], programs, or institutions involved in the testing." These include classroom assessments or quizzes, assessments for learning, or formative assessments; tests used in program evaluation or to determine unit readiness; and tests used for research purposes, including longitudinal studies to monitor change, where there are no direct consequences to the individual based on test performance.

Evidence-Centered Design

Evidence-centered design is an approach to assessment design that ensures when building the assessment that the assessment process includes collecting validity evidence. The evidence-centered design framework starts from the claims one wishes to make about a test taker or group of test takers and then works backward by considering the strength of evidence needed to support those claims and approaches for eliciting evidence relevant to the claims. It also provides specifications for analyzing and modeling domains and for implementing and delivering an assessment. In addition to supporting the design of assessments generally—from traditional to novel and innovative assessments—evidence-centered design can support the development of intelligent tutoring systems that include assessments as part of their design, including when those systems are based on simulation and game-based assessments.

CONSTRUCTS – COMPETENCIES, KNOWLEDGE, AND ATTITUDES TO BE LEARNED AND THE STATE OF ASSESSMENT

Constructs are the attributes assessed to reflect either learning that has occurred or the readiness to learn further, and as this report has demonstrated, the list of potential constructs to assess in the military context is vast. The Army's Talent Attribute Framework,[38] introduced in Chapter 2, exemplifies the wide range of knowledge, skills, and attributes that military members may need in the Army. Traditionally, constructs have been characterized as either hard skills, such as cognitive and technical competencies, or soft skills, such as teamwork, communication, and curiosity. In the upcoming section, both hard skills and soft skills are discussed. (pp. 262-265)

Contexts and Types of Training

There are several contexts of adult learning to consider for determining the appropriateness of assessments and evaluations. In formal training and education contexts, standardized tests are typically administered and accompanied by a more formal evaluation design. Training in an informal context – e.g., informal field-based learning – is not often accompanied by formal assessments or formal evaluation designs. Similarly, self-regulated or self-directed learning involves the self-teaching and self-study to create durable competencies useful for on-the-job performance and career advancement, which often does not involve formal evaluation. Another context in the military is the after-action review, which is almost never subject to assessments or evaluation designs.

There are also variations in the kinds of military training that can be administered. Technical and operational training pertain to the instruction and learning of the technical and operational competencies needed for the job. (pp. 268-69)

Performance Methods

Performance methods include approaches for drawing inferences about the level of an attribute possessed by a participant in a performance task, including standardized tests.

---

[38] *Army Talent Attribute Framework: A Unified Framework for Defining Personnel Characteristics*, Royston et al., U.S. Army Research Institute, Aug. 2022.

This is the most mature methodology for assessment, and the literature on human capabilities relies almost entirely on performance on standardized tests. Much of psychometrics – the psychology discipline devoted to testing, measurement, assessment, and related activities – is based on analyses of performance on standardized tests. Simulations are performance tasks that extend the capabilities of standardized tests and expand the ways individuals can be assessed. There is considerable research on using simulations for assessment.

One exciting development is the attempt to develop performance measures for soft 21st century skills, such as communication, collaboration, and creativity. This includes stealth tests – methods of continually tracking a person's progress while providing immediate, automated responses – to measure creativity and gamification techniques.

VALIDITY, RELIABILITY, AND FAIRNESS OF ASSESSMENTS AND EVALUATIONS

The foundations of testing are the psychometric qualities of tests, which are their validity, reliability, and fairness. Issues of the validity, reliability, and fairness of evaluations can also be addressed independently of the validity, reliability, and fairness of the assessments used in those evaluations. Although the concepts are similar, the terminology is different and the concepts come from the experimental design and causal inference literatures, not from the testing, assessment, and psychometrics literature. The following section focuses on test validity.

Test Validity

Although there are different perspectives on the validity of assessment scores, two have been particularly influential. One is the *validity argument*, which sees the validation process as one of obtaining evidence from multiple sources to draw inferences supporting the use of a test for a particular purpose for a particular population. This has come to be the mainstream view within psychometrics, as reflected in *The Standards* (AERA et al., 2014). A challenge for this view is that this concept of validity does not conform to the relatively simple idea, and the belief commonly held by researchers, that validity concerns whether a test measures what it purports to measure. Thus, in response to the *validity argument* view, an alternative conceptualization is referred to as *ontological realism*. This is the idea that a test is a valid measure of a construct to the degree that when the level of the construct changes, the test score changes along with it, just as when the temperature changes the thermometer reading does, too. A challenge for this alternative view is that psychological constructs are often fuzzy or poorly understood, unlike the case with many physical constructs such as temperature.

There has been a concern within psychology with deficiencies in constructs, which has resulted in either downplaying the role of constructs or pleas for better measurement. An example of the former is the focus on modeling specific responses to replace constructs. Examples of the latter are suggestions for upscaling operational definitions of constructs and better matching constructs and measures. There has also been a proposal to reformulate constructs, recognizing their explanatory function and emphasizing their heterogeneity and variability in content and measurement and their hierarchical nature

from general to specific. Validity generalization is an important concept in testing applications to justify the use of tests in a particular setting. It is based on the idea that the use of high-stakes testing not only is supported by local validity studies but can also be justified by combining similar studies through a meta-analysis.

General Assessment Validity Threats

Scholars have suggested that there are two overarching threats to assessment validity: construct-irrelevant variance, and construct underrepresentation. *The Standards* (AREA, 2014) defines these two concepts as follows: Construct-irrelevant variance is "variance in test-taker scores that is attributable to extraneous factors that distort the meaning of scores and thereby decrease the validity of the proposed interpretation." Construct underrepresentation is "the extent to which a test fails to capture important aspects of the construct domain that the test is intended to measure, resulting in test scores that do not fully represent that construct."

Examples of construct-irrelevant variance would be a test of mathematics proficiency that used complex English language which caused trainees, particularly non-native English speakers, to perform worse than they might otherwise perform had simpler language been used. Situational judgment tests, such as using unfamiliar scenarios to assess collaboration, might be another example of construct-irrelevant variance. An example of construct underrepresentation would be an assessment of one of the so-called Big 5 personality traits, such as conscientiousness, that failed to include items measuring one of its facets, such as rule-following. Another example would be a general reasoning test that failed to include inductive reasoning. (p. 271-73)

Validity Threats for High-Stakes Testing

When an assessment is used for high-stakes purposes, such as for personnel selection and classification, receipt of a special, desirable assignment, or (from the standpoint of a leader) unit success, then cheating is a likely validity threat. If the stakes apply to the individual, then the cheating concern would focus on the individual; if the high stakes applied to a program or class, the cheating concern would focus on the individual responsible for that program or class. Grade inflation is an example of a practice that is a response to the use of grades for high-stakes purposes. For assessments involving ratings, lack of inter-rater reliability is also a source of validity threats. This specific concern involves both high- and low-stakes assessment.

… Finally, anxiety, such as test anxiety or math anxiety, is another validity threat to high-stakes testing, as test takers might be impeded from displaying their true ability because of test-related anxiety. A meta-analysis concluded that test anxiety's effects can be in the small to moderate range, which can be significant in decision making. …

Reliability/Precision of Test Scores

Reliability "is the extent to which [test scores] are consistent across different occasions of testing, different editions of the test, or different raters scoring the test taker's responses." (p. 274)

Fairness: Disparate Effect, Measurement Invariance, and Beyond
Fairness is a complex and often fraught topic. Within testing and psychometrics, it is considered "a fundamental validity issue" (AERA et al., 2014). The basic concern as it pertains to fairness is avoiding measurement bias. *The Standards* additionally includes accessibility, meaning "an unobstructed opportunity to demonstrate [all test takers'] standing on the construct(s) being measured" and universal design, which addresses accessibility issues through all stages of test design. The Educational Testing Service's fairness guidelines define fairness as "the extent to which inferences and actions based on test scores are valid for diverse groups of test takers," and reviews construct-irrelevant barriers to success including knowledge, skill, and ability barriers such as specialized knowledge, translation, religion, unfamiliar item types, and U.S. culture, and emotional barriers…. (pp. 275-76)

Modeling Change and Competencies Growth Longitudinally
There is a large literature on modeling longitudinal data relevant to the assessment and evaluation of adult learning and developmental outcomes. These sources review issues associated with modeling change generally. For the modeling of specific competencies invoked during instruction, the literature on adaptive instructional systems, or intelligent tutoring systems is informative. These systems typically include a student model, which represents the student's current level of knowledge regarding the curricular elements, or knowledge components being taught.

Bayesian knowledge tracing models the learning of knowledge components (competencies) over time, as shown by performance on exercises, as a function of initial skill mastery, learning the skill after item completion, forgetting the skill, and making slips (answering incorrectly despite having mastered the skill), and guesses (answering correctly despite not having mastered the skill). Shen and colleagues review the considerable research base on different knowledge tracing approaches, which they classify into (a) Bayesian models (a kind of Hidden Markov Model [HMM]), (b) logistic models (learning factors analysis, performance factor analysis), or (c) deep learning models (deep knowledge tracing, memory-aware knowledge tracing, attentive knowledge tracing, graph-based knowledge tracing), along with hybrid models. The review finds that deep learning models tend to outperform Bayesian and logistic models, but at the expense of interpretability. Variants on traditional knowledge tracing approaches accommodate learner differences in learning rate or knowledge, learner engagement, forgetting, and side information (student or performance information other than responses, such as response times, whether the tutor intervenes, item text content, and student language proficiency). (pp. 277-278)

VALIDITY IN EVALUATION
Validity in evaluation pertains to the validity of inferences about the effects of a treatment on outcomes or associations of a treatment with outcomes. For example, an inference that a particular intervention led to an increase in learning outcomes might be more or less valid, depending on whether the evaluation design included a proper control group and whether the treatment was implemented properly. As with test scores, for

> which validity is not a property of the test, with evaluations validity is not a property of the method or design of the evaluation, but of the inferences drawn about the relationships between treatment and outcomes. (p. 279)

Given the fact that the only way to access educational programs that lead to high return on investment credentials is through IQ type testing regimens, the validity of IQ tests predicting socioeconomic success have never been established. That is, since those who do not meet the minimum IQ threshold for high stakes education programs – that rely on IQ type test taking talents as compared to hands-on talents – are not allowed access to the high stakes education and credentials, we can never know how the optimized test-taking group versus a control group would compare in a side-by-side analysis. Therefore, the IQ type testing regime is, in all probability, built on a house of cards since the academic establishment is largely insulated from real-world competitive forces.

In the report's conclusions, Chapter 8, it is suggested that greater investment be placed in helping military personnel discover their vocational interests. Of course, the same should take place at the secondary school levels.

> The Army Office of the Deputy Chief of Staff for Personnel should implement validated vocational interest tools to gauge military recruits' interests and leverage recruits' sense of autonomy by better aligning occupational placement to vocational interests, in accordance with mission constraints. For example, the Army could implement the Adaptive Vocational Interest Diagnostic assessment that has been evaluated in Army samples, or develop new vocational interest assessments that more closely parallel those adopted by the Navy and Air Force. (p. 305)

Evaluating Assessment Strategies

Janio and Gaff's article, "From Apprenticeship to Competency: What Anthropology Can Teach Us About Learning," (*The Evolllution*, May 28, 2025) lays out a framework for the true purpose of assessment and how best to capture competence in students – and it's not true-false or multiple choice tests, and it is not grounded in test completion times as we currently have it.

If we were to return to the methods used throughout history to determine competence in learning, current testing methodologies would be marginalized.

Janio and Gaff look to history to reevaluate assessment strategies:

> Imagine a young apprentice kneeling beside a master potter, hands coated in clay, mimicking the careful pressure and rhythm needed to shape a bowl. Across the world and throughout time, this scene has repeated itself in thousands of forms: a child shadowing a hunter, a novice doctor assisting in surgery or a philosophy student testing ideas through dialogue. Learning – real learning – has always unfolded in shared human experience, grounded in demonstration, repetition, feedback and correction.

The science of anthropology reminds us that education has long been an embodied, public practice. For most of human history, skills were not claimed but shown, and proof of learning was never a grade or credential but the competence to act.

Across cultures and centuries, learning was grounded in apprenticeship. A learner watched, listened, practiced and repeated until competence was achieved. Whether in pottery, farming, law or music, instruction was anchored in doing. The master demonstrated a skill, and the apprentice learned by performing it.

Anthropologists show that this model was never just tradition; it was a survival strategy. A poorly built shelter or incorrectly prepared food could endanger an entire community. That's why cultures developed systems of direct instruction, observation and correction. Mastery had to be visible. Competence meant survival, not symbolic recognition.

Modern education, by contrast, often relies on proxy indicators of learning: grades, diplomas, credit hours and survey data like student evaluations or satisfaction scores. But decades of research have questioned the validity of these measures, with studies showing that grades frequently reflect compliance, test-taking ability or socioeconomic background rather than actual skill acquisition. Diplomas attest to time spent, not necessarily skills learned. Student evaluations, though widely used, have been found to correlate more strongly with instructor charisma or grading leniency than with actual student learning outcomes. In short, these measures may signal participation or perception, but they do not consistently capture what a learner can do.

Behavior cuts through the noise. When students are asked to perform, produce or solve problems in real-world contexts, their learning becomes observable and verifiable. Performance-based assessment rooted in demonstrated action has repeatedly been shown to provide more accurate evidence of skill attainment and knowledge transfer. At the end of the day, we trust what's done, not what's declared. …

Reimagining Assessment

Today's classrooms often prioritize convenience over authenticity. Multiple-choice tests, timed essays and grade curves are scalable and efficient, but they rarely reflect deep or transferable learning. If a potter only memorized clay terminology, they would never make any. If a doctor could list symptoms but never diagnose or treat, they would never save any lives. Outside the classroom, promotions depend on doing the job, not passing a test.

However, many educational systems cling to indirect indicators. These shortcuts persist because they're easy to standardize and quantify, but the trade-off is high. They often reward surface learning and disadvantage students who lack cultural or linguistic fluency with academic norms.

Behavior-based assessment offers a fairer and more inclusive approach. When we evaluate what students do, not just what they say, we create more pathways to success. Learners from a wide range of backgrounds can demonstrate their knowledge through

action. The measure becomes competence, not conformity, and the advantage shifts to those who can perform, not just those who test well. …

Returning to What Works
Reclaiming assessment practices rooted in observable behavior is not a return to some nostalgic past. It is a reorientation toward what has always worked: competence through demonstration. In an era of growing skepticism about higher education's value, this shift is also a path toward credibility. When we show what students can do and those skills are relevant and real, we may restore confidence in education.

Anthropology reminds us how humans have always learned: through observation, repetition, correction and performance. Behaviorism sharpens that view by showing us that learning is visible in action. Together, they challenge us to build assessment systems rooted not in recall but in demonstration. What if courses were designed around what students must do, not just what they must know? What if final exams became performance projects, solutions or real-world tasks? What if grades reflected skills in motion, not memories on paper?

We already rely on performance in the world beyond school: to certify a pilot, evaluate a surgeon or hire a carpenter. It's time our classrooms caught up. If we want to assess learning honestly, we must begin where learning has always lived – in behavior – not because it's simple but because it's true.

This article reveals the consequences of our shift from traditional education methodologies to contemporary bureaucratic, "scientific" strategies that have sought social efficiencies to accomplish social engineering ends. The authors might be surprised to learn that the social engineers (who embraced the German statist model[39] around the turn of the last century) who designed the "efficient" education system, weren't concerned about deeper learning and transfer. Their goal was to indoctrinate students to be molded into workers who could fit neatly into workspaces that society needed to be filled. This required work environments that were also engineered to reduce the need for skills so that just about anyone could do the job (Taylorism was the science to achieve this end).

While the influence of statism has been profound in the U.S., and around the world for that matter, Americans have been sharply divided on this political movement. It has ebbed and flowed as its negative consequences have surfaced. It appears the relatively new movement called, 21st Century Skills, is marginalizing the statist bureaucratic models and shifting toward individualism (not narcissism) to help individuals follow the idea Maslow espoused with his *hierarchy of needs*.

The National Research Council (NRC) provides extensive reports on how to develop 21st century skills in students. The committee's report, *Knowing What Students Know: The Science and Design of Educational Assessment* (2001), provides an analysis of public education assessment

[39] "The principle or policy of concentrating extensive economic, political, and related controls in the state at the cost of individual liberty." *Source*: dictionary.com

research that is focused on the college-for-all model, so it has its limitations. However, it offers a great deal of insight into the development of 21st century skills. We'll start off with their summary of the purposes for assessment:

> From teachers' informal quizzes to nationally administered standardized tests, assessments have long been an integral part of the educational process. Educational assessments assist teachers, students, and parents in determining how well students are learning. They help teachers understand how to adapt instruction on the basis of evidence of student learning. (p. 19)

While the report is packed with highly insightful analyses, the committee was charged with "improving educational assessment in general and assessment of science and mathematics education in particular." (p. 17) While these two subjects are very important, there is so much more to education that must be accomplished. Therefore, we must consider what the report offers, and see where their findings might fit in an entirely new paradigm of education that sees the college track as one of a large number of choices.

Since the purpose of current public education is to prepare students for a college journey, the assessment design has been strictly based on an academic orientation. Of course, there are exceptions where career technical education (CTE) is gaining some ground (such as in Indiana), though these are rare. But even where CTE is gaining ground, non-CTE classes in those schools, which are the lion's share of requirements, remain academically oriented rather than technically oriented. So even a career-oriented curriculum is not supported with technically rigorous subjects that provide deep learning and transferable knowledge to the real-world. Therefore, a CTE track is a misnomer: CTE courses, which there are typically few to choose from, are treated not unlike a college minor where only a very limited number of courses are allowed to be taken, with the rest of the high school requirements being made up of academic, college prep courses that are largely useless to the CTE student.

A New Approach is Recommended

Chapter One of the report begins by pointing out that the current educational system is lacking in assessing literacy and numeracy competence in a coherent manner. Contemporary research literature offers greater understanding of cognition, learning, and assessment that needs to be implemented in current education to achieve competence in what assessment should do. Recent research reveals improved methodologies to help many more students achieve competence in academic subjects with the use of assessment as a tool to achieve this, rather than using it exclusively to audit memory retention and as a means for postsecondary grooming and selectivity. However, the committee admits it is unlikely that the research findings will be adopted in classrooms any time soon.

The report states: "The key issues that emerge from the themes discussed [in this report] strongly suggest that it is appropriate and necessary to rethink the scientific principles and philosophical assumptions that serve as the foundations of educational assessment." (p. 33)

A great deal of progress has been made as it relates to deeper learning and transfer of learning for academically oriented purposes, but academic subjects are only a portion of what individuals

need to know to thrive in a complex and vibrant society such as ours – and academic knowledge is certainly not the most important information citizens need to know. Academic knowledge is foundational; that is, it provides the means to learn what is important to know. Because of this, the academic community believes academic subjects should be mastered before anything else. This is a fundamental flaw in their thinking. Academic subjects should be taught in tandem with applied, contextual knowledge and should be seen as secondary, i.e., supportive, of applied learning. Without application, academic knowledge is largely an accumulation of disconnected data and procedures that are floating in the ether just waiting for a purpose.

The NRC committee explains the purpose of this report:

> This report addresses assessments used in both classroom and largescale contexts for three broad purposes: to assist learning, to measure individual achievement, and to evaluate programs. The purpose of an assessment determines priorities, and the context of use imposes constraints on the design. Thus it is essential to recognize that one type of assessment does not fit all. (p. 2)

In the Executive Summary, the committee summarizes the purpose of assessment from a public perspective:

> Educational assessment seeks to determine how well students are learning and is an integral part of the quest for improved education. It provides feedback to students, educators, parents, policy makers, and the public about the effectiveness of educational services. With the movement over the past two decades toward setting challenging academic standards and measuring students' progress in meeting those standards, educational assessment is playing a greater role in decision making than ever before.

While the first portion of this paragraph is what most of us would agree is a sensible summary, the last sentence reflects the troublesome bias in the U.S. system that drives public education toward the college-for-all cliff. "Academic standards" rule the educational landscape, with applied and technically oriented literacy receiving no attention.

To better understand the difference, think of technical writing contrasted with creative writing; or applied math contrasted with pure math; or applied science contrasted with pure science. The applied subjects are more appropriate for the vast majority of students while the abstract subjects are better suited to a small percentage of the school age population who are destined for graduate work (approximately 4.5% of the population). This is made evident when we analyze the number of jobs in *pure* science-, math-, and language-oriented occupations in society (*see* O*NET OnLine to look at job requirements). Yet, our public system has been ruled by the theoretically oriented academics since the 1894 recommendations of The Committee of Ten who established core curricula subjects that were designed for college preparation but required all students to take. The Committee claimed that a college prep curriculum was appropriate for occupational preparation as well, but this was a falsehood they were fully aware of given the momentum of the manual and industrial arts movement of the period (*see* Anderson, *History of Manual and Industrial School Education*, 1926). This has been a fundamental flaw with the system ever since and is yet to be corrected for.

Context of Assessments
The committee identifies four broad themes in discussing the context of assessments:

- "Any assessment is based on three interconnected elements or foundations: the aspects of achievement that are to be assessed (cognition), the tasks used to collect evidence about students' achievement (observation), and the methods used to analyze the evidence resulting from the tasks (interpretation)." *See* "The Assessment Triangle" below, Figure 2-1. For more in-depth analysis of the Assessment Triangle, *see* "The Assessment Triangle" section of the report, pages 44-51.

- Determine what individuals should know in our society, which determines what must be taught and assessed to ensure competency.

- "Existing assessments are the product of prior theories of learning and measurement. … [They have] contributed to … limitations and impeded progress in assessment design." (p. 19)

- Significantly improved assessment methodologies are now available for adoption in order to improve student learning.

  Every educational assessment, whether used in the classroom or large-scale context, is based on a set of scientific principles and philosophical assumptions, or *foundations* as they are termed in this report. First, every assessment is grounded in a conception or theory about how people learn, what they know, and how knowledge and understanding progress over time. Second, each assessment embodies certain assumptions about which kinds of observations, or tasks, are most likely to elicit demonstrations of important knowledge and skills from students. Third, every assessment is premised on certain assumptions about how best to interpret the evidence from the observations to draw meaningful inferences about what students know and can do. (p. 20)

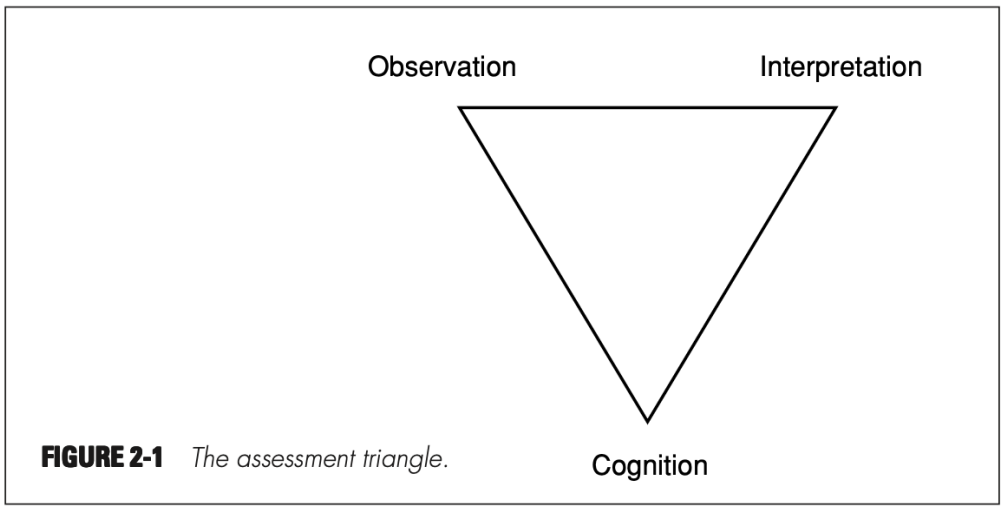


FIGURE 2-1 *The assessment triangle.*

Traditional Assessment

The committee points out that the foundational assumptions that have accompanied 20th century testing methods create "conflicts about the meaning and value of assessment results." (p. 21)

As to the third point mentioned above, it is certain that an assessment based on the 20th century multiple choice/true-false type tests will require very different interpretation methods compared to demonstrations of real-world, applied type testing. 20th century versus applied scenarios are incomparable since 20th century testing typically measures memory and recall abilities of disassociated facts and procedures; whereas applied assessments are self-evident and therefore reflect competence and proficiency which requires far less interpretation and is therefore more objective. After all, if subject matter is abstract, assessments must also be in the abstract realm, which makes interpretation more complicated – other than verbatim recitation.

The report sums up the historical flaws with the assessment paradigm:

> Current assessments are … derived from early theories that characterize learning as a step-by-step accumulation of facts, procedures, definitions, and other discrete bits of knowledge and skill. Thus, the assessments tend to include items of factual and procedural knowledge that are relatively circumscribed in content and format….
>
> The most common kinds of educational tests do a reasonable job with certain functions of testing, such as measuring knowledge of basic facts and procedures and producing overall estimates of proficiency for an area of the curriculum. But both their strengths and limitations are a product of their adherence to theories of learning and measurement that fail to capture the breadth and richness of knowledge and cognition. The limitations of these theories also compromise the usefulness of the assessments. The growing reliance on tests for making important decisions and for improving educational outcomes has called attention to some of their more serious limitations.
>
> … Traditional [i.e., 20th century] tests do not focus on many aspects of cognition that research indicates are important, and they are not structured to capture critical differences in students' levels of understanding.
>
> The limits on the kinds of competencies currently being assessed also raise questions about the validity of the inferences one can draw from the results. If scores go up on a test that measures a relatively narrow range of knowledge and skills, does that mean student learning has improved, or has instruction simply adapted to a constrained set of outcomes? If there is explicit "teaching to the test," at what cost do such gains in test scores accrue relative to acquiring other aspects of knowledge and skill that are valued in today's society? This is a point of considerable current controversy.
>
> A second issue concerns the usefulness of current assessments for improving teaching and learning – the ultimate goal of education reforms. On the whole, most current large-scale tests provide very limited information that teachers and educational administrators can use to identify why students do not perform well or to modify the conditions of instruction in ways likely to improve student achievement. The most widely used state

and district assessments provide only general information about where a student stands relative to peers (for example, that the student scored at the 45th percentile) or whether the student has performed poorly or well in certain domains (for example, that the student performs "below basic in mathematics"). Such tests do not reveal whether students are using misguided strategies to solve problems or fail to understand key concepts within the subject matter being tested. They do not show whether a student is advancing toward competence or is stuck at a partial understanding of a topic that could seriously impede future learning. Indeed, it is entirely possible that a student could answer certain types of test questions correctly and still lack the most basic understanding of the situation being tested, as a teacher would quickly learn by asking the student to explain the answer (see Box 1-2). In short, many current assessments do not offer strong clues as to the types of educational interventions that would improve learners' performance, or even provide information on precisely where the students' strengths and weaknesses lie.

BOX 1-2 Rethinking the Best Ways to Assess Competence

Consider the following two assessment situations:

Assessment #1

Question: What was the date of the battle of the Spanish Armada?
Answer: 1588 [correct].
Question: What can you tell me about what this meant?
Answer: Not much. It was one of the dates I memorized for the exam. Want to hear the others?

Assessment #2

Question: What was the date of the battle of the Spanish Armada?
Answer: It must have been around 1590.
Question: Why do you say that?
Answer: I know the English began to settle in Virginia just after 1600, not sure of the exact date. They wouldn't have dared start overseas explorations if Spain still had control of the seas. It would take a little while to get expeditions organized, so England must have gained naval supremacy somewhere in the late 1500s.

Most people would agree that the second student showed a better understanding of the Age of Colonization than the first, but too many examinations would assign the first student a better score. When assessing knowledge, one needs to understand how the student connects pieces of knowledge to one another. Once this is known, the teacher may want to improve the connections, showing the student how to expand his or her knowledge.

> A third limitation relates to the static nature of many current assessments. Most assessments provide "snapshots" of achievement at particular points in time, but they do not capture the progression of students' conceptual understanding over time, which is at the heart of learning. This limitation exists largely because most current modes of assessment lack an underlying theoretical framework of how student understanding in a content domain develops over the course of instruction, and predominant measurement methods are not designed to capture such growth. (pp. 26-28)

> … While concerns associated with large-scale tests have received considerable attention, … the classroom assessments commonly used by teachers also are often limited in the information they provide. Just as large-scale tests have relied on an incomplete set of ideas about learning, so, too, have the kinds of assessments teachers regularly administer in their classrooms. Often, teachers adhere to assessment formats and scoring practices found in large-scale tests. This can be traced largely to teacher education programs and professional development experiences that have for the most part failed to equip teachers with contemporary knowledge about learning and assessment, especially the knowledge needed to develop tasks that would elicit students' thinking skills or make it possible to assess their growth and progress toward competence. (p. 29)

Alternative approaches to assessment have been developing. Personally, I feel that "performance assessments" offer the most significant advances since they are tied to real-world scenarios. Project- and Problem-Based Learning strategies are excellent alternative examples of learning and assessment. The report sums it up:

> The quest for alternatives to traditional assessment modes has led many states to pursue approaches that include the use of more open-ended tasks that call upon students to apply their knowledge and skills to create a product or solve a problem. Performance assessment represents one such effort to address some of the limitations of traditional assessments. Performance assessment, an enduring concept … requires students to perform more "authentic" tasks that involve the application of combined knowledge and skills in the context of an actual project. (p. 30)

The committee sums up the usefulness of assessments:

> Every assessment, regardless of its purpose, rests on three pillars: a model of how students represent knowledge and develop competence in the subject domain, tasks or situations that allow one to observe students' performance, and an interpretation method for drawing inferences from the performance evidence thus obtained. In the context of large-scale assessment, the interpretation method is usually a statistical model that characterizes expected data patterns, given varying levels of student competence. In less formal classroom assessment, the interpretation is often made by the teacher using an intuitive or qualitative rather than formal statistical model. (p. 2)

Let's break the "pillars" down:

1. *A model of how students represent knowledge and develop competence in the subject domain*: In traditional education methodologies, students' representation of knowledge is in disconnected bits of data or facts; with competence being the ability to recall the data for shallow, close transfer uses. In deep learning, transferable, and technical education, performance needs to be in applied settings which determines competence; that is, the ability to accomplish a useful task that has value in the real world.
2. *Tasks or situations that allow one to observe students' performance*: But what is the purpose of the observation? If it is simply to see if content data has been retained and then recalled quickly and accurately, but without a sense of purpose in application, the knowledge is then void of substance. However, if the purpose of the observation is to determine competence in useful activities, "performance" then has significance worthy of assessing.
3. *An interpretation method for drawing inferences from the performance evidence thus obtained*: If interpretation is to infer statistical data for a given population, this may be helpful for administrators, policymakers, and researchers, but this has marginal use for individuals, who should be the primary beneficiaries of assessment.

In the end, "Improvements in learning will depend on how well assessment, curriculum, and instruction are aligned and reinforce a common set of learning goals, and on whether instruction shifts in response to the information gained from assessments." (p. 25)

The Ability to Reason in Contrast with the Ability to Memorize
The committee opens the dialogue in considering what should be taught in schools:

> Expectations about what all students should learn – and, by implication, what they should be tested on – have changed in response to social, economic, and technological changes and as a result of the standards-based reform movement. All students are now expected to demonstrate the kinds of reasoning and problem-solving abilities once expected of only a minority of young people. Assessments are needed to gauge these aspects of student competence. (p. 33)

Stanovich (*What Intelligence Tests Miss*, 2009) explains how the education system misses the mark on instruction for developing reasoning abilities.

Large-scale Versus Classroom Assessment Purposes
The report attempts to reconcile large-scale assessment with assessing individual competence, which, historically, have not been complimentary in purpose. It is important to understand the motives and purposes of these forms of assessment in educational programs. The committee explains:

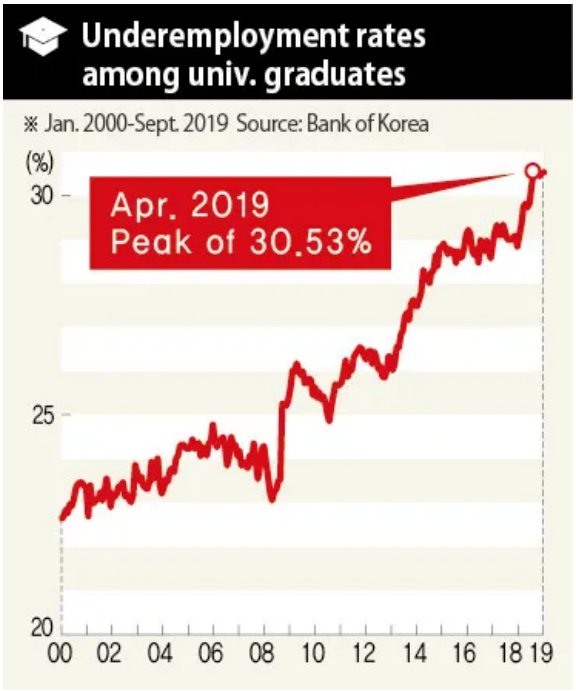
> The power of classroom assessment resides in its close connections to instruction and teachers' knowledge of their students' instructional histories. Large-scale, standardized assessments can communicate across time and place, but by so constraining the content and timeliness of the message, they often have limited utility in the classroom. Thus the contrast between classroom and large-scale assessments arises from the different purposes they serve and contexts in which they are used. Certain trade-offs are an inescapable aspect of assessment design.

Students will learn more if instruction and assessment are integrally related. In the classroom, providing students with information about particular qualities of their work and about what they can do to improve is crucial for maximizing learning. It is in the context of classroom assessment that theories of cognition and learning can be particularly helpful by providing a picture of intermediary states of student understanding on the pathway from novice to competent performer in a subject domain.

… Whereas teaching directly to the items on a test is not desirable, teaching to the theory of cognition and learning that underlies an assessment can provide positive direction for instruction.

To derive real benefits from the merger of cognitive and measurement theory in large-scale assessment, it will be necessary to devise ways of covering a broad range of competencies and capturing rich information about the nature of student understanding. Indeed, to fully capitalize on the new foundations described in this report will require substantial changes in the way large-scale assessment is approached and relaxation of some of the constraints that currently drive large-scale assessment practices.

… The current educational assessment environment in the United States assigns much greater value and credibility to external, large-scale assessments of individuals and programs than to classroom assessment designed to assist learning. … More of the research, development, and training investment must be shifted toward the classroom, where teaching and learning occur. (pp. 8-9)



Underemployment rates among Korean univ. graduates.

See "30% of S. Korean graduates overqualified for their current jobs," *Hankyoreh*:

Parents and instructors need to understand that bureaucracies support government interests, not individual interests. They must therefore understand that State and Federal demands for assessing our children may have indirect benefits to determine if educational trends are heading in the right direction, but the immediate concern of parents is their children's needs and success, and to subordinate State interests. This becomes even more relevant when we look at the intent of educational bureaucrats whose primary goal is to see everyone graduate college regardless of individual or societal need since college provides status to bureaucracies. But Korea's mistake of pushing college-for-all proves this strategy to be illogical.

Design Considerations of Assessment

With curriculum designed for deep learning and for transfer, the authors suggest a new approach to instruction and assessment is required.

> A model of cognition and learning should serve as the cornerstone of the assessment design process. This model should be based on the best available understanding of how students represent knowledge and develop competence in the domain.
>
> The model of learning can serve as a unifying element – a nucleus that brings cohesion to curriculum, instruction, and assessment. This cohesive function is a crucial one because educational assessment does not exist in isolation, but must be aligned with curriculum and instruction if it is to support learning. (p. 3)

However, if we are to structure education based on something as effective and socially beneficial as the Swiss system, a large-scale cohesive model of assessment becomes complicated after primary education is complete. This is due to the multitude of occupational training paths that must be provided for in the secondary years (*see* BLS's Occupational Outlook Handbook to see just how broad the spectrum for occupations is). The further up the pyramid of educational requirements one climbs for a given occupation, the more specialized it becomes. Assessment must match the specialization that occupations are composed of. Large-scale assessments must then take a 30,000-foot view of society. Data driven academic content must then be replaced with an understanding of what makes a society function from a big picture perspective: familially, legally, politically, philosophically, theologically, economically, psychologically, sociologically, etc. We might say that all these disciplines fall under civics understanding since civics is all about membership in a community in all its manifestations.

The committee points to the importance of designing assessment and the appropriate interpretation of what needs to be assessed. The committee suggests the following to determine competency in a task:

> Starting with hypotheses about the cognitive demands of a task, a variety of research techniques, such as interviews, having students think aloud as they work problems, and analysis of errors, can be used to analyze the mental processes of examinees during task performance. Conducting such analyses early in the assessment development process can help ensure that assessments do, in fact, measure what they are intended to measure. (p. 7)

In Chapter 1, under the section titled "An Era of Higher Standards and High-Stakes Tests" (p. 23), the report addresses "challenging academic standards" assembled by professional associations of subject matter specialists who have developed standards outlining content knowledge that public schools should teach in the principle subject disciplines.

> Many of the standards developed by states, school districts, and professional groups emphasize that it is important for students not only to attain a deep understanding of the content of various subjects, but also to develop the sophisticated thinking skills necessary to perform competently in these disciplines. … Several of the standards also stress the need for students to build coherent structures of knowledge and be able to apply that knowledge in much the same manner as people who work in a particular discipline. (p. 23)

This portion of the report reveals a serious bias in the halls of academia: Public education is dedicated to the handful of academic subjects taught in schools, with the full expectation – albeit, subconsciously – that students will pursue careers in one of these academic subjects. All other occupations are ignored. This is grounds for the complete abandonment of the secondary school years unless a student's trajectory is to pursue one of the academic subjects being offered, or a career that is closely aligned.

The professional associations mentioned above were listed as:

- Mathematics standards have been developed by the National Council of Teachers of Mathematics
- The science standards developed by the NRC
- The standards in several subjects developed by New Standards.
    - New Standards Performance Standards Vol. 1: Elementary School – 4th grade
    - New Standards Performance Standards Vol. 2: Middle School – 8th grade
    - New Standards Performance Standards Vol. 3: High School – 10th grade

Given the bias just discussed, it is certain that these standards are dedicated to the academic subjects with the intent of preparing students for careers in one of these subjects. Certainly there is a need to determine their value as it relates to the *relative worth* principle. Equally important to analyze is how the standards could be distinguished between academic pursuits and non-academic trajectories. Every occupation has different requirements, but academia has tended to ignore these differences, making their jobs easier by subjectively determining what they perceive as important rather than doing in-depth research into every occupation to determine what every occupational requirement will be for the novice. This is unjust to absolutely every student who is expected to learn subject matter that is useless to them. Individuals' time is precious and should not be taken for granted.[40]

This raises questions about the idea of standards. While it is important to establish foundational education the primary years are meant to cover, standards for the secondary years require serious

[40] But if acculturation into institutional life is the goal, with the importance of subject matter being secondary, then a bloated curriculum is useful to extend the time in school to complete the assimilation process, which Progressives of the early part of the last century estimated to be twelve years to accomplish.

revision. O*NET OnLine offers an important resource for customizing educational standards for the various occupational families that make up the economy. Under "Find Occupations," see the various categories to get an idea of a focus the various career categories require from education:

Career Cluster: The National Career Clusters® Framework serves as an organizing tool for programs, curriculum design and instruction. Career Clusters contain occupations in the same field of work that require similar skills. They can help focus plans towards obtaining the necessary knowledge, competencies, and training for success in a particular career pathway.

Industry: Industries are broad groups of businesses or organizations with similar activities, products, or services. Occupations are included based on the percentage of workers employed in that industry.

Job Family: Job Families are groups of occupations based on work performed, skills, education, training, and credentials.

Job Zone: Job Zones group occupations into one of five categories based on levels of education, experience, and training necessary to perform the occupation.

STEM: These occupations are listed that require education in science, technology, engineering, and mathematics disciplines.

All Occupations: There are 1,016 occupation titles and codes within the current O*NET system.

Another important resource to further investigate clusters of economic activity to design educational standards is the Bureau of Labor Statistics' Occupational Outlook Handbook.

Armed with the ability to analyze occupational family trajectories, a student's education program can be customized to the economic sector the individual is destined for. A general education in tandem with a relatively focused and sequentially graduated training regime can provide an individual a broad education that avoids entrapment by a narrow set of skills and knowledge that vocational education was previously accused of. This provides the needs for citizenship plus a path for lucrative career outcomes.

State Control vs. Local Autonomy

The report reveals conflicts within a public system that home- and micro-schooling are immune to:

> Decisions about assessment, curriculum, and instruction are further complicated by actions taken at different levels of the educational system, including the classroom, the school or district, and the state. Each of these levels has different needs, and each uses assessment data in varied ways for somewhat different purposes. Each also plays a role in making decisions and setting policies for assessment, curriculum, and instruction, although the locus of power shifts depending on the type of decision involved. Some of these actions emanate from the top down, while others arise from the bottom up. States generally exert considerable influence over curriculum, while classroom teachers have

> more latitude in instruction. States tend to determine policies on assessment for program evaluation, while teachers have greater control over assessment for learning. This situation means that adjustments must continually be made among assessment, curriculum, and instruction not only horizontally, within the same level (such as within school districts), but also vertically across levels. For example, a change in state curriculum policy will require adjustments in assessment and instruction at all levels. (p. 52)

This is a moving target where teachers shift priorities as political forces swing randomly. This is not healthy for students to be exposed to, but most importantly, it is clear that the needs of individuals are at the bottom of this totem pole.

Theoretical Underpinnings of Cognitive Research

The next chapter, Chapter 3, "Advances in the Sciences of Thinking and Learning" refers to NRC's report *How People Learn: Bridging Research & Practice* (1999).

This chapter explains the growth of cognitive psychology over the course of the 20th century and highlights some of the theoretical underpinnings of the research. It was a clumsy progression that eventually coalesced into a rational approach in understanding how people learn. The report sums up 4 theories that developed and eventually coalesced. The four are: *The differential perspective*, *the behaviorist perspective*, *the cognitive perspective*, and *the situative perspective*. (pp. 60-63)

The *cognitive perspective* is summed up nicely with the following:

> In cognitive theory, knowing means more than the accumulation of factual information and routine procedures; it means being able to integrate knowledge, skills, and procedures in ways that are useful for interpreting situations and solving problems. Thus, instruction should not emphasize basic information and skills as ends in themselves, but as resources for more meaningful activities. As Wiggins (1989) points out, children learn soccer not just by practicing dribbling, passing, and shooting, but also by actually playing in soccer games.
>
> Whereas [other] approaches focus on how much knowledge someone has, cognitive theory also emphasizes what type of knowledge someone has. An important purpose of assessment is not only to determine what people know, but also to assess how, when, and whether they use what they know. This information is difficult to capture in traditional tests, which typically focus on how many items examinees answer correctly or incorrectly, with no information being provided about how they derive those answers or how well they understand the underlying concepts. Assessment of cognitive structures and reasoning processes generally requires more complex tasks that reveal information about thinking patterns, reasoning strategies, and growth in understanding over time. (pp. 62-63)

Utilizing the cognitive perspective is the approach I refer to as employing abstract, theoretical knowledge to support applied learning – i.e., the pure academic disciplines supporting the

applied disciplines in a secondary role, in contrast to being the primary educational material. In this way, deeper learning and the ability to transfer are realized.

Next, the report addresses what is referred to as the *situative perspective*:

> Instead of viewing thought as individual response to task structures and goals, the situative perspective describes behavior at a different level of analysis, one oriented toward practical activity and context. Context refers to engagement in particular forms of practice…. (p. 63)
>
> Most current testing practices are not a good match with the situative perspective. Traditional testing presents abstract situations, removed from the actual contexts in which people typically use the knowledge being tested. From a situative perspective, there is no reason to expect that people's performance in the abstract testing situation adequately reflects how well they would participate in organized, cumulative activities that may hold greater meaning for them.
>
> From the situative standpoint, assessment means observing and analyzing how students use knowledge, skills, and processes to participate in the real work of a community. For example, to assess performance in mathematics, one might look at how productively students find and use information resources; how clearly they formulate and support arguments and hypotheses; how well they initiate, explain, and discuss in a group…. (p. 64)

A good analogy might be one of business financial decisions an owner is frequently faced with. On an abstract test measuring data and procedural competence, a person's responses to typical academically oriented questions might be graded as poor, but in a real-world business scenario, the poor test taker may outperform the straight A student every time. In business, context is critical. This ability cannot be captured in the traditional test-taking paradigms.

The committee points out how the cognitive and situative perspectives can be complimentary:

> The cognitive perspective informs the design and development of tasks to promote conceptual development for particular elements of knowledge, whereas the situative perspective informs a view of the larger purposes and practices in which these elements will come to participate. Likewise, the cognitive perspective can help teachers focus on the conceptual structures and modes of reasoning a student still needs to develop, while the situative perspective can aid them in organizing fruitful participatory activities and classroom discourse to support that learning.
>
> Both perspectives imply that assessment practices need to move beyond the focus on individual skills and discrete bits of knowledge that characterizes the earlier associative and behavioral perspectives. They must expand to encompass issues involving the organization and processing of knowledge, including participatory practices that support knowing and understanding and the embedding of knowledge in social contexts. (p. 65)

Long-term vs. Short-term Memory

The report discusses long-term and short-term memory differences (*see* pages 3-4). I will include an important aspect of their relationship this report addresses:

> Unlike working memory, long-term memory is, for all practical purposes, an effectively limitless store of information. It therefore makes sense to try to move the burden of problem solving from working to long-term memory. What matters most in learning situations is not the capacity of working memory – although that is a factor in speed of processing – but how well one can evoke the knowledge stored in long-term memory and use it to reason efficiently about information and problems in the present. (p. 68)
>
> It can be argued that estimates of what people have stored in long-term memory and how they have organized that information are likely to be more important than estimates of working memory capacity in most instances of educational assessment. (p. 71)

I'll leave it to the reader to investigate this subject in the report. A final point needing to be made here is that public education ignores long-term memory and focuses exclusively on short-term memory, which is extremely limited for purposes of deeper understanding and transfer.

Schemas

Next, the report addresses *schemas* under the section "Schemas and the Organization of Knowledge."

> Schemas help move the burden of thinking from working memory to long-term memory. They enable competent performers to recognize situations as instances of problems they already know how to solve; to represent such problems accurately, according to their meaning and underlying principles; and to know which strategies to use to solve them. (p. 70)

The committee offers an example of comparative schemas:

> … Lopez, Atran, Coley, and Medin (1997) showed that different cultures differentially encourage certain types of schemas. In reasoning about animals, American college students tend to resort to taxonomic schemas, in which animals are related by dint of possessing common features (and indirectly, having certain genetic relationships). In contrast, the Itzaj Maya, a jungle-dwelling group in Guatemala, are more likely to reason by emphasizing ecological relationships. It is not that the Americans are unaware of ecological relations or the Maya are unaware of feature possession. Rather, each group has adopted its own schema for generalizing from an observed characteristic of one animal to a presumed characteristic of another. In each case, however, the schema has particular value for the individuals operating within a given culture.
>
> Extensive research shows that the ways students represent the information given in a mathematics or science problem or in a text that they read depends on the organization of their existing knowledge. As learning occurs, increasingly well-structured and qualitatively different organizations of knowledge develop. These structures enable

> individuals to build a representation or mental model that guides problem solution and further learning, avoid trial-and-error solution strategies, and formulate analogies and draw inferences that readily result in new learning and effective problem solving. The impact of schematic knowledge is powerfully demonstrated by research on the nature of expertise…. (p. 71)

Structuring curriculum to help students build schemas would be a very effective pedagogical strategy for deep learning and transfer as well as moving short-term knowledge to long-term storage, which should improve assessment results as long as assessments are designed to capture deeper learning abilities rather than rattle off facts and procedures.

Expert vs. Novice Ability to Retrieve Knowledge

A nice explanation of the difference between experts and novices is summarized as follows:

> What distinguishes expert from novice performers is not simply general mental abilities, such as memory or fluid intelligence, or general problem-solving strategies. Experts have acquired extensive stores of knowledge and skill in a particular domain. But perhaps most significant, their minds have organized this knowledge in ways that make it more retrievable and useful. (p. 72)

> Most important, they have efficiently coded and organized this information into well-connected schemas. … Of particular interest to researchers is the way experts encode, or chunk, information into meaningful units based on common underlying features or functions. Doing so effectively moves the burden of thought from the limited capacity of working memory to long-term memory. (p. 73)

The following point made by the committee demonstrates why instruction in disconnected facts and procedures is not conducive to learning:

> The knowledge that experts have cannot be reduced to sets of isolated facts or propositions. Rather, their knowledge has been encoded in a way that closely links it with the contexts and conditions for its use.

> These and other related findings suggest that teachers should place more emphasis on the conditions for applying the facts or procedures being taught, and that assessment should address whether students know when, where, and how to use their knowledge. (p. 73)

The fact that experts closely link knowledge with contexts and conditions of use, determines how we learn, how we remember, and how we use knowledge. An applied education system will serve this revelation beautifully. The traditional education paradigm can now be seen for what it is: a contrived institutional system that serves the institution but not individuals.

Metacognition

Metacognition can be expressed as “thinking about thinking.”

> Experts have strong metacognitive skills (Hatano, 1990). They monitor their problem solving, question limitations in their knowledge.… This capability for self-regulation and self-instruction enables advanced learners to profit a great deal from work and practice by themselves and in group efforts. (p. 78)

The committee asserts that metacognition is one of the most important aspects of cognition – "the process of reflecting on and directing one's own thinking."

> Experts use metacognitive strategies for monitoring understanding during problem solving and for performing self-correction. Assessment of knowledge and skill in any given academic domain should therefore attempt to determine whether an individual has good metacognitive skills. (p. 79)

Teaching and reinforcing metacognitive skills must take precedence over the memorization of data and procedures. As a matter of fact, data and procedures – the stuff contemporary tests are made of – should be used as the tools to reinforce such learning skills as building resourcefulness, developing schemas, and practicing metacognitive strategies. The same principle applies to practical application of knowledge in contrast to studying abstract theories. A primary goal for education is to prepare individuals so that they may educate themselves and not be dependent upon anyone other than themselves to learn whatever they wish or need to learn.

Transfer

The committee turns to the transfer of learning. They reference research that reveals the problem with connecting abstractions with application. This helps explain the limitations of the traditional bureaucratic assessment culture. That is: If abstractions are all that is known, transferring that knowledge to application will prove extremely difficult. The means to do well on tests then, must rely on regurgitation of what was memorized. While this ability is considered appropriate to be considered "college material," it is not sufficient to be considered career ready since such employees will have to be micro-managed. Employers want employees to be able to hit the ground running, and hopefully be able to teach the employer a thing or two with new insights.

> Insights about learning and transfer have come from studies of situations in which people have failed to use information that, in some sense, they are known to have. Bassok and Holyoak (1989) showed, for example, that physics students who had studied the use of certain mathematical forms in the context of physics did not recognize that the same equations could be applied to solve problems in economics.[41] On the other hand, mathematics students who had studied the same mathematical forms in several different contexts, but not economics, could apply the equations to economics problems.[42] (p. 87)

If we take a serious look at this, it makes perfect sense that it would be difficult to transfer without understanding fundamentals. For example: Though a mathematician immersed in logic can create computations for different scenarios that have similar underlying structures (example of *far transfer*), the student, who only memorized computations, would find it difficult, if not

[41] This is due to the silo effect of subject disciplines.
[42] This demonstrates that transfer across domains is possible.

impossible, to create computations for scenarios outside the context in which a computation was learned (this defines *close transfer*). Without an understanding of math's logic, students are merely clerks doing computations – like Mr. Cratchit working for Scrooge in *A Christmas Carol*. Cratchit was a "slave" of Scrooge's given his severe limitations. Scrooge's connectiveness to the real world provided the source of value for wealth and job creation – his cruelty and immorality aside.

This point is further reinforced where the report provides:

> Learners must develop an understanding of when (under what conditions) it is appropriate to apply what they have learned. Recognition plays an important role here. Indeed, one of the major differences between novices and experts is that experts can recognize novel situations as minor variants of situations for which they already know how to apply strong methods. Transfer is also more likely to occur when the person understands the underlying principles of what was learned. (p. 87)

Indeed! What is explained here is a logical understanding of the application of the required mathematical knowledge; which the expert possesses but the novice does not. Scenarios may change dramatically (e.g., physics versus economics), but the mathematical logic can remain unchanged, which is evident to the expert. This helps explain why novices are unable to transfer knowledge between subject domains in school – something NRC has accepted as true in their *How People Learn* reports. Without foundational understandings rooted in logic, domains become silos with no bridges to connect them.

Bassok and Holyoak's (1989) research demonstrates that if domains are not isolated (i.e., siloed), transfer across domains is common. Unlike academia, industry proves this to be true on a regular basis.

Applied versus Abstract Knowledge

A study done on literacy practices exemplifies a point I make throughout my various works: Applied knowledge is superior to purely abstract knowledge; and applied knowledge is readily available to all. The following study demonstrates this:

> Hull, Jury, Ziv, and Schultz (1994) studied literacy practices in an electronics assembly plant where work teams were responsible for evaluating and representing their own performance. Although team members had varying fluency in English, the researchers observed that all members actively participated in the evaluation and representation processes, and used texts and graphs to assess and represent their accomplishments. This situation suggests that reading, writing, quantitative reasoning, and other cognitive abilities are strongly integrated in most environments, rather than being separated into discrete aspects of knowledge. Tests that provide separate scores may therefore be inadequate for capturing some kinds of integrated abilities that people need and use on the job. (p. 89)

The academic types would score better on tests compared to non-academic type of people, but it doesn't mean they can transfer test-taking talents to real-world applications. Translation:

Academic test-taking abilities do not translate to superior occupational abilities. However, since education is optimized for academic test-taking abilities, many lucrative credentials are guarded by academia. The relief valve for this protectionism is business acumen, manufacturing, and the apprenticed professions. These can be highly lucrative but without academia's controls.

The chapter concludes with

> Studies of expert-novice differences in subject domains illuminate critical features of proficiency that should be the targets for assessment. … One of the most important roles for assessment is the provision of timely and informative feedback to students during instruction and learning so that their practice of a skill and its subsequent acquisition will be effective and efficient. … Assessments of academic achievement need to consider carefully the knowledge and skills required to understand and answer a question or solve a problem, including the context in which it is presented, and whether an assessment task or situation is functioning as a test of near, far, or zero transfer. (p. 103)

This is very different from the traditional memory, recall, and speed assessment priorities.

Design With Students in Mind

The committee provides a summary of what the first chapter of Part III covers. It lays out the approach for instructors to judge whether an assessment is designed with students as the priority.

> Five key features of a new approach to assessment design serve as the organizing themes for this chapter:
>
> • A model of cognition and learning, or a description of how people represent knowledge and develop competence …, is a cornerstone of the assessment development enterprise. Unfortunately, the model of learning is not made explicit in most assessment development efforts, is not empirically derived, and/or is impoverished relative to what it could be.
>
> • To increase the chances of collecting evidence that supports the types of inferences one wants to draw, the design and selection of assessment tasks, along with the procedures for evaluating students' responses, should be guided jointly by the cognition and interpretation elements of the assessment triangle [i.e., cognition, observation, and interpretation]. An assessment should be more than a collection of tasks that work well individually. The utility of assessment information can be enhanced by considering how to design and/or select tasks so that the information derived from them can be combined to support the desired inferences.
>
> • The process of construct validation during test design should rest, in part, on evidence that tasks actually tap the cognitive content and processes intended.
>
> • Although reporting of results occurs at the end of an assessment cycle, assessments must be designed from the beginning to ensure that reporting of the desired types of information will be possible and effective. When results are reported, well-delineated

descriptions of learning in the domain are key to their effectiveness for communicating about student performance.

• Fairness in testing is defined in many ways, but at its core is the idea of comparable validity: a fair assessment is one that yields comparably valid inferences from person to person and group to group. One way of thinking about fairness is to take into account individual learners' instructional histories when designing an assessment. (p. 176)

Additionally, fairness must take into account the different proclivities and talents individuals possess. For example: The talent of possessing a photographic memory cannot be used as the bar that all other students must live up to. This would lead to the optimization of the education system for those with such a talent, while leaving all others to find alternatives to formal education. While the current public system is not quite this extreme (otherwise the pool would be too small for the system to function), there is a definite bias for those who have excellent memories, regardless of reasoning abilities.

Limitations of Assessing People

The committee points out that "the features of assessment design [described in the report] represent an ideal case that is unlikely to be fully attained with any single assessment." Given all the proclivities and talents people possess – and the countless walks of life society is composed of – it would be impossible for assessments to capture what encompasses the full array of talents and competencies needed in a complex society. Indeed, since pedagogical approaches fall far short of the needs of individuals and society, assessing individual's competencies in preparation for entering a social community must also be understood to be inadequate to achieve the ideal. The assessment of domains – i.e., subject matter of individual disciplines – must be seen as assessing foundational understandings that allow individuals to learn what they will need in order to become competent in an occupation and competent for full participation in a constitutionally organized democratic republic based on free market (i.e., *invisible hand*) principles. Full competency/mastery can only take place through consistent application once foundations have been laid. Therefore, educational assessments in subject domains should be seen as rough estimates of preparation for career training.

Career training allows for discovering individual native talents to determine suitability for a chosen occupation. Its purpose can be associated with Boot Camp (Navy) or Basic Training (Army & Air Force), or Recruit Training (Marines) that helps determine suitability for military duty. However, recruit training has an important purpose of testing individuals' wherewithal to endure intense combat situations. While civilian occupational training does not rise to such high stakes considerations, the correlation is worth noting since some occupations do encompass life-death scenarios – e.g., think of architects (e.g., safety of structures), civil engineers (e.g., safety of bridges), mechanics (e.g., think of our dependency on vehicular braking systems), physicians, firemen, police, etc. No one person has all, or even most of the talents that make up the full array of occupational competencies – and IQ type tests don't even scratch the surface for revealing abilities required in the real world since no occupational duties and responsibilities require test-taking skills.

However, IQ type tests may suggest a secondary attribute that is useful for determining success: conscientiousness, which is highly associated with success. That is, those who work hard to perform well on assessments may be displaying high levels of conscientiousness rather than "intelligence" in the manner in which we think of "intelligence." That is, conscientious students study hard to do well on tests in the same manner they assert themselves to do well on any endeavor they choose to pursue. The *act* of studying hard, therefore, should not necessarily be seen as closely correlated with the *gift* of memory and recall, which people think of as "intelligence." This is an association that is rarely considered in the literature; though Poropat did analyze the correlation.

Nevertheless, one must be careful to avoid assigning a direct correlation between conscientiousness and IQ type test results. There are many highly conscientious-minded individuals who see IQ type tests as irrelevant to their lives since academia is not an attractive world to them.

Identifying the Stepping Stones of Learning

The section *The Importance of a Model of Cognition and Learning* touches on the complexity of deciding what to assess. "Existing guidelines for assessment design emphasize that the process should begin with a statement of the purpose for the assessment and a definition of the content domain to be measured (AERA et al., 1999; Millman and Greene, 1993)." (p. 178)

> It is not, however, the goal of education to make all school children experts in every subject area, and many would argue that "literacy" and "competency" are more appropriate goals. Ideally, then, a model of learning will also provide a developmental perspective, laying out one or more typical progressions from novice levels toward competence and then expertise, identifying milestones or landmark performances along the way. The model of learning might also describe the types of experiences that provoke change or learning. Models of learning for some content areas will depict children as starting out with little or no knowledge in the domain and through instruction gradually building a larger and larger knowledge base.
>
> … There is no single way in which knowledge is represented by competent performers, and there is no single path to competence. But some paths are traveled more than others. When large samples of learners are studied, a few predominant patterns tend to emerge. (p. 182)
>
> Differences among learners should not be ignored. Thus a third key feature of a model of learning is that it should convey a variety of typical ways in which children come to understand the subject matter of interest.
>
> … Fourth, starting with a theory of how people learn the subject matter of interest, the designers of an assessment will need to select a slice or subset of the larger theory of cognition and learning as the assessment targets. That is, any given model of learning underlying an assessment will capture some, but not all, aspects of what is known about how students think and learn in the domain. That selection should depend on the purpose for the assessment.

Finally, a model of learning will ideally lend itself to being aggregated in a principled way so that it can be used for different assessment purposes. (p. 184)

[When developing assessments,] basic principles of cognition and learning described in Chapter 3 – such as the importance of how people organize knowledge, represent problems, and monitor their own learning – can inform the translation of curriculum into instruction and assessment. The principle that learning must start with what students currently understand and know about a topic and build from there will always hold. (p. 185)

Facets of Student Thinking

The Facets program provides an example of how student performance can be described at a medium level of detail that emphasizes the progression or development toward competence and is highly useful for classroom assessment. Developed through collaboration between an experienced high school science teacher and a cognitive psychologist, the assessment approach is based on models of learning termed *facets of student thinking*. The approach is predicated on the cognitive principle that students come to instruction with initial ideas and preconceptions that the teacher should identify and build on.

The term facets refers to pieces of knowledge or reasoning, processes, beliefs, or constructions of pieces of knowledge that serve as a convenient unit of thought for analysis of student behavior. In many ways they behave like general rules that students have in their knowledge base about how the world works. Facets are derived from research and from teachers' observations of student learning. (pp. 186-87)

Progress Maps

Progress maps provide a description of skills, understandings, and knowledge in the sequence in which they typically develop – a picture of what it means to improve over time in an area of learning.

… These maps can serve as the basis for assessments for both large-scale and classroom purposes. "Progress is monitored in much the same way as a child's physical growth is monitored: from time to time an estimate is made of a student's location on a developmental continuum, and changes in location provide measures of growth over time" (Masters and Forster, 1996, p. 1). (p. 190)

Task Design

Messick (1994) distinguishes between a task-centered and a construct-centered approach to assessment design…. With a task-centered approach, the focus is on having students perform meaningful and important tasks, and the target of inference is, implicitly, the tendency to do well on those tasks. Such an approach makes sense under certain circumstances, such as an arts contest or figure-skating competition, when evaluation of the product or performance per se is the focus. But educational decision makers are rarely concerned with one particular performance. They tend to be more interested in the

> underlying competencies that enable performance on a task, as well as on a range of related activities. In such cases, a construct-centered approach is needed. Such an approach starts with identifying the knowledge, skills, or other attributes that should be assessed (expressed through the model of learning), which then guide the selection or construction of relevant tasks and scoring procedures.
>
> … A related point is that design should focus on the cognitive demands of tasks (the mental processes required for successful performance), rather than primarily on surface features, such as how tasks are presented to students or the format in which students are asked to respond. For instance, it is commonly believed that multiple-choice items are limited to assessing low-level processes such as recall of facts, whereas performance tasks elicit more complex cognitive processes. However, the relationship between item format and cognitive demands is not so straightforward. Although multiple-choice items are often used to measure low-level skills, a variety of item formats, including carefully constructed multiple-choice questions, can in fact tap complex cognitive processes. Similarly, performance tasks, usually intended to assess higher-level cognitive processes, may inadvertently tap low-level ones. (p. 194)
>
> Ideally, task difficulty should be explained in terms of the underlying knowledge and cognitive processes required, rather than simply in terms of statistical item difficulty indices, such as the proportion of respondents answering the item correctly. Beyond knowing that 80 percent of students answered a particular item incorrectly, it would be educationally useful to know why so many did so, that is, to identify the sources of the difficulty so they could be remedied (assuming, of course, that they represented important construct-relevant variance). Cognitive theory and analysis can be helpful here. (p. 195)

The committee reveals a flaw psychometricians miss when designing assessments:

> The most commonly used statistical measures of task difficulty ignore the fact that two tasks with similar surface features can be equally difficult, but for different reasons. For example, two language tasks that require combined reading and writing skills may exhibit the same overall level of difficulty according to the statistical item parameters, even though one task places greater demands on compositional skills and the other on reading comprehension. (p. 196)

Such oversight is no small matter when the stakes can be very high for individuals. But then, psychometricians rationalize their shortcomings by referring to averages of a population. Translation: No one is special other than those who score high on the subjective academic tests they created. Further, Todd Rose, in his book *The End of Average*, has revealed the evils of relying on averages. Statistics may be useful to understand trends in entire populations or cohorts, but their use for judging individuals is evil. After all, individuals are unique given all the talents and desires they possess that differ from every other individual; whereas statistics has to ignore this fact. Statistics are used to rank individuals academically based on a very narrow set of parameters that don't transfer to the real world very well. Yet, students' entire futures are dictated by the results of their statistical rank. Statistically based assessments, when used to rank individuals, are at war with individuals.

Quality vs. Quantity

Contrary to current educational fashions, the report points out that the quality of understanding typically is more important than quantity of information. Voluminous knowledge is worthless if it is not available in the mind for use in the real world. This requires associations be made in the mind. Appropriate and sufficient associations define the quality of learning.

> Like tasks, scoring methods must be carefully constructed to be sensitive to critical differences in levels and types of student understanding identified by the model of learning. At times one may be interested in the quantity of facts a student has learned, for instance, when one is measuring mastery of the alphabet or multiplication table. However, a cognitive approach generally implies that when evaluating students' responses, the focus should be on the quality or nature of their understanding, rather than simply the quantity of information produced. (p. 197)

Elaborating on the "alphabet" analogy, perhaps a distinction between early primary school learners – where the quantify might be the most important measurement – and later primary school learners where quality of students' understanding can be expanded upon. Late primary school learners can be introduced to the evolution of written language, explaining how, originally, cuneiforms were used but eventually were replaced by alphabets. Wikipedia explains the evolution as follows:

> Beginning in the 9th century BC, adaptations of the Phoenician alphabet thrived, including Greek, Old Italic and Anatolian scripts. The alphabet's attractive innovation was its phonetic nature, in which one sound was represented by one symbol, which meant only a few dozen symbols to learn. The other scripts of the time, cuneiform and Egyptian hieroglyphs, employed many complex characters and required long professional training to achieve proficiency; which had restricted literacy to a small elite.

Differences Between Students

The following is an example of open-mindedness and a recognition that psychometricians have been a serious problem in assessing students for a very long time. Their techniques have been flawed (and unfortunately, are now baked in the cake) and it's time we recognize that the educational landscape has been far more of the country's problem than student learning and comprehension abilities. The one-size-fits-all college prep approach, along with its marriage to statistics, is evil. Customization for many walk-of-life choices is necessary and just.

> Silver, Alacaci, and Stylianou (2000) have demonstrated some limitations of scoring methods used by the National Assessment of Educational Progress (NAEP) for capturing the complexities of learning. They reanalyzed a sample of written responses to an NAEP item that asked students to compare two geometric figures and found important differences in the quality of the reasoning demonstrated: some students showed surface-level reasoning (paying attention to the appearance of the figures), others showed analytic reasoning (paying attention to geometric features), and still others demonstrated more sophisticated reasoning (looking at class membership). Despite these qualitative differences, however, the NAEP report simply indicated that 11 percent of students gave satisfactory or better responses – defined as providing at least two reasons why the shapes

> were alike or different – while revealing little about the nature of the students' understanding. Whereas the current simple NAEP scoring strategy makes it relatively easy to control variation among raters who are scoring students' responses[43], much other information that could have educational value is lost. (p. 200)

The written responses allowed for the observation of differences in the quality of reasoning which is challenging, if not impossible, to achieve with multiple choice questions. Written or verbal explanations were the historic and traditional assessment methods used for millennia to determine depth of learning.

Task Sets and Assembly of an Assessment Instrument

In the section of the report, the committee contrasts flaws found in traditional math assessment methods compared to a nuanced cognitive approach that seeks identification of student shortcomings in order to make corrections:

> An assessment should be more than a collection of items that work well individually. The utility of assessment information can be enhanced by carefully selecting tasks and combining the information from those tasks to provide evidence about the nature of student understanding. Sets of tasks should be constructed and selected to discriminate among different levels and kinds of understanding that are identified in the model of learning. To illustrate this point simply, it takes more than one item or a collection of unrelated items to diagnose a procedural error in subtraction. If a student answers three of five separate subtraction questions incorrectly, one can infer only that the student is using some faulty process(es), but a carefully crafted collection of items can be designed to pinpoint the limited concepts or flawed rules the student is using. (p. 200)

QUASAR

The report covers the math instructional program named QUASAR.

> QUASAR[44] is an instructional program developed by Silver and colleagues to improve mathematics instruction for middle school students in economically disadvantaged communities (Silver, Smith, and Nelson, 1995; Silver and Stein, 1996). To evaluate the impact of this program, which emphasizes the abilities to solve problems, reason, and communicate mathematically, assessments were needed that would tap the complex cognitive processes targeted by instruction. In response to this need, the QUASAR Cognitive Assessment Instrument was developed (Lane, 1993; Silver and Lane, 1993).
>
> Assessment design began with the development of a model of learning. Using the Curriculum and Evaluation Standards for School Mathematics … the assessment developers specified a number of cognitive processes important to competent performance in the domain: understanding and representing problems; discerning mathematical relations; organizing information; using and discovering strategies,

[43] This is the reason assessments have been designed in this manner ever since bureaucratic control sought social efficiency goals and whose primary purpose was to train students to be "good soldiers" rather than free-thinkers.
[44] Cai et al. (1996) work demonstrated the effectiveness of the QUASAR program in improving students' thinking and reasoning.

> heuristics, and procedures; formulating conjectures; evaluating the reasonableness of answers; generalizing results; justifying an answer or procedures; and communicating. These processes were defined more specifically in each of the content categories covered by the assessment: number and operations; estimation; patterns; algebra, geometry, and measurement; and data analysis, probability, and statistics. (p. 209)

Reporting on Assessment Results

Norm-referenced tests are achievement tests whose purpose is to compare students' performance relative to other students who take the same test.

> Comparative test information can help parents, teachers, and others determine whether students are progressing at the same rate as their peers or whether they are above or below the average. Norm-referenced data are limited, however, because they do not show what a student actually can or cannot do. A score indicating that a student is in the 40th percentile in mathematics does not reveal what mathematics knowledge the student does or does not have. The student may have answered most items correctly if the norm group was high-performing, or may have answered many questions incorrectly if the norm group performed less well. Nor does the norm-referenced score indicate what a student needs to do to improve. (pp. 212-14)

This information may be useful to education administrators and policy makers, but students and their parents should be very cautious about interpreting this information for their own use. The question must be asked: Is the test material even relevant to the student? For example: If the student's occupational trajectory requires no algebra, does poor performance on algebra – and therefore the effect it has on percentile standing – even matter? The same question holds for science. Therefore, if science and algebra are not of any use to the student's future, why should their inclusion in tests, hold so much sway over the future of students who will have no need of them? Given the fact that most people will graduate high school and work for private companies, why are we not testing them for business and technical knowledge that will be useful to them and their employers.

Given the twisted bias of public secondary school curriculum and the associated biased testing regime, parents may want to think twice about participation unless it serves their children in some fashion. Since students should be provided custom education, assessments should be customized as well. After all, the only thing that matters are individual students – certainly not the educational bureaucracies whose sole purpose is supposed to prepare individuals for life.

Lexicon vs. Vernacular Usage in Instruction

The following statement by the committee reveals a desire to embrace educating all children rather than just those who are academically talented:

> It would … be educationally useful to analyze the difficulty of an assessment task in terms of which students get it wrong, and why it is so problematic for those students. Part of the answer might lie in differences in the communicative practices students bring to the assessment. (p. 196)

Indeed, every human endeavor has its own lexicon[45], not to mention regional and cultural nuanced vernacular. If a test-taker is not proficient in terminology or phrases used in an assessment, the results are not likely to be good – in the same manner as if a test-taker were not acquainted with the material covered in a test. In both cases, "intelligence" is not being tested. Understanding the question is the first step in being able to answer a question appropriately. Then, the understanding of the material and competency in its application can be measured. This will require subject specialists to take a serious look at their lexicon to determine if words and/or phrases need to be simplified into the vernacular[46], or invest more time in ensuring students are very familiar with the subject's lexicon.

This gets to the issue of fairness. Fairness may have different meanings to different people, but one thing is certain, fairness implies something will have a similar positive, negative, or neutral effect upon all people involved. Think of law: The law must be applied and enforced upon **all** citizens equally.

The committee makes an important argument where it points out:

> An assessment task is considered biased if construct-irrelevant characteristics of the task result in different meanings for different subgroups. For example, it is now common wisdom that a task used to observe mathematical reasoning should include words and expressions in general use and not those associated with particular cultures or regions; the latter might result in a lack of comparable score meanings across groups of examinees. (p. 215)

In addition to the *particular cultures or regions'* general use of words and expressions referenced above, we should add to this principle the use of words that are discipline-specific that can bias tests. Effort should be invested into rephrasing the nomenclature of the various subject disciplines into the vernacular that the laity easily comprehends. For those on occupational trajectories destined for a specialty at the secondary and/or postsecondary levels, a shift to formal terminology – i.e., the lexicon of a discipline – for their occupations should be transitioned into. Since the nomenclature across the multitude of occupations possess unique lexicon, it is not possible to teach all the "languages" of each to the entire school age population, so why do we attempt to do so for the core academic subjects exclusively when the vast majority of students are not destined to specialize in these subjects?

Assessment in Practice

> Although assessments are currently used for many purposes in the educational system, a premise of this report is that their effectiveness and utility must ultimately be judged by the extent to which they promote student learning. The aim of assessment should be to educate and improve student performance, not merely to audit it [which is all that is currently occurring].

---

[45] In this case, the meaning of "lexicon" "is the vocabulary of a … branch of knowledge (such as nautical or medical)" that is specific to that branch but typically is not used outside the branch. *Source* Wikipedia

[46] In this case, the meaning of "vernacular" is "the ordinary, informal spoken form of language" in contrast to "institutionally promoted, literary, or formal" language. *Source*, Wikipedia

> … Assessments do not function in isolation; an assessment's effectiveness in improving learning depends on its relationships to curriculum and instruction. Ideally, instruction is faithful and effective in relation to curriculum, and assessment reflects curriculum in such a way that it reinforces the best practices in instruction. (p. 221)

Rest assured that the connectedness between current curriculum, instruction, and assessment does not progress in a coherent whole throughout the primary and secondary public-school years. They tend to be haphazard and poorly connected, which challenges deeper understanding and transfer abilities. When instruction is haphazard, society must rely on individuals' unique gifts in a subject domain to eventually fully acquire competence and eventual mastery. Special gifts allow for individuals to make connections that are otherwise out of the reach to most people. Whereas, if subject material and instruction are well organized, a majority of people can acquire competence and if committed to, eventually acquire mastery.

Classroom Assessment

> The focus in this section of the report is on assessments used by teachers to support instruction and learning, also referred to as *formative assessment*. Such assessment offers considerable potential for improving student learning…. (p. 225)
>
> In response to research evidence that simply giving grades on written work can be counterproductive for learning (Butler, 1988), teachers began instead to concentrate on providing comments without grades – feedback designed to guide students' further learning. Students also took part in self-assessment and peer-assessment activities, which required that they understand the goals for learning and the criteria for quality that applied to their work. … In these ways, assessment situations became opportunities for learning, rather than activities divorced from learning. (p. 227)
>
> There is a rich literature on how classroom assessment can be designed and used to improve instruction and learning (e.g., Falk, 2000; Niyogi, 1995; Shepard, 2000; Stiggins, 1997; Wiggins, 1998). This literature presents powerful ideas and practical advice to assist teachers across the K-16 spectrum in improving their classroom assessment practices.
>
> … At the 2000 annual meeting of the American Educational Research Association, Shepard (2000) began her presidential address by quoting Graue's (1993, p. 291) observation, that "assessment and instruction are often conceived as curiously separate in both time and purpose." Shepard asked: How might the culture of classrooms be shifted so that students no longer feign competence or work to perform well on the test as an end separate from real learning? Could we create a learning culture where students and teachers would have a shared expectation that finding out what makes sense and what doesn't is a joint and worthwhile project, essential to taking the next steps in learning? … How should what we do in classrooms be changed so that students and teachers look to assessment as a source of insight and help instead of its being the occasion for meting out reward and punishments. To accomplish this kind of transformation, we have to make

> assessment more useful, more helpful in learning, and at the same time change the social meaning of evaluation. (p. 228)

This would require instructors have a vested interest in students' success. An indifferent view on their part can lead to poor student outcomes; though the use of bell curves – i.e., the use of averages – in student outcomes provides cover to take refuge in. In other words, the bell curve provides protection from criticism since it can be used to argue that there will always be those in a population who fail. While logically this may sound reasonable, morally, this bakes failures into the system. For example: In some circles of academia, there are professors who are of the opinion that if a certain percentage of students do not fail, the class isn't "rigorous" enough. These people should not be teaching. While there may be students who will fail for a variety of reasons, it should be the hope and expectation of instructors that they manage to get every student across the competency finish line if it's within their power.

Computer Tutors

> [I]ntelligent tutoring systems are powerful examples of the use of cognitively based classroom assessment tools blended with instruction. … Students make mistakes, and the system offers effective remediation. As a result, students on average learn more with the system than with other, traditional instruction.
>
> … In a study of Anderson's geometry tutor with high school students and their teachers, Schofield and colleagues found that teachers provided more articulate and better-tuned feedback than did the intelligent tutor (Schofield, Eurich-Fulcer, and Britt, 1994). Nevertheless, students preferred tutor-based to traditional instruction, not for the reasons one might expect, but because the tutor helped teachers tune their assistance to problems signaled by a student's interaction with the tutor. Thus, student interactions with the tutor (and sometimes their problems with it) served to elicit and inform more knowledgeable teacher assistance, an outcome that students apparently appreciated. Moreover, the assistance provided by teachers to students was less public. Hence, formative assessment and subsequent modification of instruction – both highly valued by these high school students – were mediated by a triadic relationship among teacher, student, and intelligent tutor. (pp. 231-33)

With artificial intelligent tutors, this should provide for flexible educational programs that will take instruction to the next level.

Collaboration versus Competition

Under the section "The Quality of Feedback," the committee reports on the unhealthy culture of competition in academia. Yet another reason parents should strongly consider home- or micro-schooling for their children. In addition, teachers in the public arena should seriously consider opening their own micro-schools.

> The culture of focusing on grades and rewards and of seeing classroom learning as a competition appears to be deeply entrenched and difficult to change. This situation is more apparent in the United States than in some other countries (Hattie, Biggs, and Purdie, 1996). The competitive culture of many classrooms and schools can be an

> obstacle to learning, especially when linked to beliefs in the fixed nature of ability. Such beliefs on the part of educators can lead both to the labeling – overtly or covertly – of students as "bright" or "dull" and to the confirmation and enhancement of such labels through tracking practices.
>
> International comparative studies – notably case studies and video studies conducted for the Third International Mathematics and Science Study that compare mathematics classrooms in Germany, Japan, and the United States – highlight the effects of these cultural beliefs. The studies underscore the difference between the culture of belief in Japan that the whole class can and should succeed through collaborative effort and the culture of belief endemic to many western countries, particularly the United States, that emphasizes the value of competition and differentiation (Cnen and Stevenson, 1995; Holloway, 1988).
>
> The issues involved in students' views of themselves as learners may be understood at a more profound level by regarding the classroom as a community of practice in which the relationships formed and roles adopted between teacher and students and among students help to form and interact with each member's sense of personal identity (Cobb et al., 1991; Greeno and The Middle-School Mathematics Through Applications Group, 1997). Feedback can either promote or undermine the student's sense of identity as a potentially effective learner. For example, a student might generate a conjecture that was later falsified. One possible form of feedback would emphasize that the conjecture was wrong. A teacher might, instead, emphasize the disciplinary value of formulating conjectures and the fruitful mathematics that often follows from generating evidence about a claim, even (and sometimes especially) a false one.
>
> A voluminous research literature addresses characteristics of learners that relate to issues of feedback. … The important point to be made here is that teachers should be aware that different types of feedback have motivational implications that affect how students respond. Black and Wiliam (1998) sum up the evidence on feedback as follows: ". . . the way in which formative information is conveyed to a student, and the context of classroom culture and beliefs about ability and effort within which feedback is interpreted by the individual recipient, can affect these personal features for good or ill. The hopeful message is that innovations which have paid careful attention to these features have produced significant gains when compared with the existing norms of classroom practice." (235-36)

The above summary provides evidence that teacher and peer biases can dramatically influence an individual's chances of success. This is a gamble, and one that cannot be justified under present cultural forces. This warrants a shift to home- or micro-schooling where such negative cultural influences can be monitored and avoided, which supports the concept of using school vouchers. In addition, "intelligent tutors" offer an unbiased approach to learning since computers should not have such biases programmed into them.

Fairness

Cognitive research has shown that students classified as less able learners have demonstrated the largest learning gains when formative assessment methods (such as ThinkerTools) are incorporated in the learning process. This reveals the shortcomings of traditional methodologies and that there are different learning styles for different types of people. The fact that this hasn't been known – or admitted – reveals just how unfair the public education system has been for well over one hundred years.

Continuity

> In addition to comprehensiveness and coherence, an ideal assessment system would be designed to be continuous. That is, assessments should measure student progress over time, akin more to a videotape record than to the snapshots provided by the current system of on-demand tests. To provide such pictures of progress, multiple sets of observations over time must be linked conceptually so that change can be observed and interpreted. Models of student progression in learning should underlie the assessment system, and tests should be designed to provide information that maps back to the progression. With such a system, we would move from "one-shot" testing situations and cross-sectional approaches for defining student performance toward an approach that focused on the processes of learning and an individual's progress through that process (Wilson and Sloane, 2000). Thus, continuity calls for alignment along the third dimension of time. (pp. 256-57)

Text Analysis and Scoring

Historical assessment methods used in classrooms, before bureaucratic State systems took control, did not use multiple choice tests. Students were given assignments or asked questions which they were expected to articulate answers to or explanations of the task assigned. Teachers were able to determine the depth of understanding students possessed. The Progressives, for numerous reasons, abandoned these age-old methods and adopted "the scientific" assessment system that was designed to efficiently – though not effectively – measure the masses for placement in society.

Computer tutoring systems have provided researchers the opportunity to discover that the age-old systems were actually superior to the "efficient" bureaucratic system that was initiated in the late 19th century and become fully adopted throughout the 20th century. The NRC committee could see the benefits and reported:

> Extended written responses are often an excellent means of determining how well someone has understood certain concepts and can express their interrelationships. In large-scale assessment contexts, the process of reading and scoring such written products can be problematic because it is so time- and labor-intensive, even after raters have been given extensive training on standardized scoring methods. Technology tools have been developed to aid in this process by automatically scoring a variety of extended written products, such as essays. (p. 269)

Written work is the best means of determining students' competence and abilities on a given assignment. Before the takeover by the bureaucrat state with its simplistic measurement of memory of data and procedures, teachers relied extensively on students' written work. With the advent of AI tutoring, a return to student written explanations is feasible once again.

Part IV Conclusion

> In the future envisioned by the committee … Teachers will assess students' understanding frequently in the classroom to provide them with feedback and determine next steps for instruction. Their classroom practices will be grounded in principles of how students think and learn in content domains and of assessment as a process of reasoning from evidence. Teachers will use this knowledge to design assessments that provide students with feedback about particular qualities of their work and what they can do to improve.
>
> Students will provide evidence of their understanding and thinking in a variety of ways – by responding to teachers' questions, writing or producing projects, working with computerized tutoring systems, or attempting to explain concepts to other students. Teachers, in turn, will use this information to modify instruction for the class and for individuals on the basis of their understanding and thinking patterns. (p. 292)

The findings in this report are not likely to be transferred to the classroom, and the committee recognizes this where it explains: "It is unlikely that the insights gained from current or new knowledge about cognition, learning, and measurement will be sufficient by themselves to bring about transformations in assessment such as those described in this report." This is due, in large part, to the size of the educational establishment along with the vested interests in maintaining the status quo. Therefore, home- and micro-schooling offer the best hope for adopting many of the recommendations in this report.

Conclusion

Strong (2017) offers a nice bullet-pointed summary of the current assessment paradigm in contrast with the way it should be pursued, which reveals why transfer of learning, in large part, does not occur due to distorted and disconnected motivations:

Reality of the Present:	**Hopes for the Future:**
* Students "learn" exclusively for a grade	* Students learn for gain in knowledge, skills and self-improvement
* Students complete work because it is worth 10 or 20 or 150 points	* Students create extremely high quality work because the work they do tells a story of who they are
* Students view teachers as authoritarians who judge students' rote content knowledge with scores	* Students are viewed as collaborators within their learning journey
* Students form their identity exclusively around grades, particularly quizzes and tests	* Students have a broader scope through which to form their academic identity

Regarding the interpretation and use of The National Assessment of Educational Progress (NAEP) achievement levels, Edley and Koenig (2017) reveal that not only is there no consensus on what constitutes proper levels of literacy and numeracy, but even the results of assessments to determine the levels of competency are not understood.

> Originally, NAEP was designed to measure and report what students in the United States *actually* know and be able to do. … Achievement levels were designed to lay out what students *should* know and be able to do. Reporting NAEP results as the percentage of students who score at each achievement level was intended to make NAEP results more understandable. This type of reporting was designed to clearly and succinctly highlight the extent to which U.S. students are meeting the nation's expectations.
>
> For this aspect of its evaluation, the committee looked for information on both the intended and the actual uses and users of NAEP achievement-level reports. We also considered the extent to which research evidence supports these uses.
>
> … Conclusion: The National Assessment of Educational Progress achievement levels are widely disseminated to and used by many audiences, but the interpretive guidance about the meaning and appropriate uses of those levels provided to users is inconsistent and piecemeal. Without appropriate guidance, misuses are likely.
>
> … The current achievement-level descriptors may not provide users with enough information about what students at a given level know and can do. The descriptors do not clearly provide accurate and specific information about the things that students at the cut score for each level know and can do. (pp. 240-42)

As in so many cases, bureaucracies are inefficient and ineffective agencies that frequently do harm to the public interest. This is why small government is so **very** important.

Much needs to be done to discover how to assess students at various stages of development to help steer them in directions that fit their talents. The one-size-fits-all college prep program is a failure in every sense of the word since it serves approximately 15 to 20% of the population – as measured by IQ related data – with another 10 to 15% receiving marginal benefits, speaking in very rough terms. But even so, it serves these populations inadequately when we consider how poorly graduates are able to transfer what they've learned and how poorly businesses see the preparedness of graduates.

The development of individual portfolios by students combined with GPAs can help demonstrate individual potential whether for college entrance or for employers. Stansbury (2016) offers a nice summary in an article covering this subject. In addition, Mintz (Jan. 2017) provides another approach. Also, Mangan (2015) offers the potential problems employers might face with portfolios with too much information. AAC&U (2017) offers this as a portfolio type of assessment system:

> [T]he VALUE [Valid Assessment of Learning in Undergraduate Education] initiative breaks new ground by basing its assessment of student learning achievement on the actual work that

> students produce in response to assignments from the formal instructional curriculum in whatever institution(s) the student attended. Rather than a standardized test divorced from the curriculum, VALUE draws evidence from the actual courses and teachers at an institution, assessing the learning artifacts (papers and assignments) produced by students to demonstrate their achievement of specific learning outcomes. (p. 3)

The ability to reason is far superior to memorization, but with our current fixation on true-false and multiple-choice type testing, reasoning abilities are incapable of being evaluated. While assessments that utilize essay writing offer a superior assessment strategy,[47] the argument against essays rests upon the problem with the evaluator's subjective judgments. But current testing methodologies are most assuredly subjective, but from a bureaucratic institutional platform, which supposedly gives it real "authority."

AAC&U's proposal to use student's actual work (papers and assignments) to evaluate their level of achievement is referred to as students' *transcripts*, which represents GPAs and test results. However, I think the word *portfolio* is a more appropriate description since it is the actual work of students rather than a summary or judgment of their work.

Walsh (2017) reflects on work currently being done by the University of Texas called ChainScript based transcript services "that can ease the process with which credentials are stored and communicated:"

> TEx allows the learner's academic record to transfer from a traditional transcript into a blockchain-powered ChainScript that can be kept by the individual learner. This ChainScript contains a validated and immutable record of the learner's academic and professional accomplishments across multiple institutions and experiences, building a portfolio of accomplishments that includes credits, competencies, micro-certificates, degrees, and other records of achievement.
>
> As frameworks like Connecting Credentials develop, and blockchain-based transcripts become more commonplace, a foundation will be in place that can do much more than just ease the process (and cost) of achieving a degree.

While the *portfolio* concept is being considered, it is believed that some sort of ranking must be assigned to the individual in order to compare each child to other children. Tests must be designed so winners and losers can be divided – i.e. cognitive sorting – with the winners receiving the lion's share of public resources in order to elevate them at the expense of all others. This is diametrically opposed to our system of government and such exclusive reliance on these methods of assessment must end on legal as well as ethical grounds. Instead of comparing children, we should be seeking to discover every individual's talents. If we developed various assessment methods based on the multiple intelligence theory, we could discover varied and vast untapped talent pools that have previously been ignored. With this, students can be guided down an appropriate path suited to their abilities and interests with both individuals and society reaping incredible rewards from it.

Guskey offers words of wisdom in tackling the effort of change:

---

[47] See Hoffmann, 2003, pp. 121-22 where he references Zahner.

> The stakes in this battle are high. They involve the professional integrity of teachers and school leaders, as well as the academic success and well-being of students.
>
> Those with the boldness to question and defy these long-held traditions are unlikely to be the most popular or well-liked leaders in their schools or districts. They might even be labeled as radicals, agitators, or troublemakers. But true leadership in education isn't about being popular or well liked. It's not about maintaining the status quo, mindlessly following traditions, or just getting by. It's not about justifying antiquated policies and practices with the simple rationalization "We've always done it that way."
>
> Rather, true leadership is about doing what's right to make a positive and lasting difference in the lives of young people. It means taking a stand on what is best for students instead of thoughtlessly adhering to traditions that have long outlived their usefulness. Sometimes that means stirring things up, asking hard questions, pushing for change, and as a result, sacrificing one's popularity. That's where the courage required to be a true educational leader comes in. (p. 8)

It's time the monopolistic power of the Statist educational system be dismantled. This means that decentralization, like Finland, is an absolute must for change to be realized. It will require the disassembling of accreditation institutions, assessment methodologies, the credentialing monopoly, and the dominant public education system. These institutions have proven themselves failures at a catastrophic level when we disassemble educational statistics. ***Any*** system would be better than the one we now suffer under. Therefore, control must be localized at the municipal, district, or county levels, but with State financial support. The contribution of the Federal government should be limited exclusively to the gathering of information, its organization, and then its dissemination to all who desire it, as was the original charter written for the National Center for Education Statistics (NCES) in 1867 (see Sparks, 2017b). No educational laws, regulations, or funding should be provided by the Federal government since it has proven to be too dangerous to be trusted. This would then justify the elimination of the Department of Education since it would no longer serve the factional interests of those in power.

As the title of this essay points out, assessment testing is not a measure of intelligence, but certainly is a means of marginalizing political participation and economic opportunity for the majority. This is an utterly unacceptable situation and a paradigm shift in educational offerings and delivery systems, correlated with appropriate assessment methods, **must** take place.

Final Thoughts

One thing this report should bring to everyone's attention, as it relates to the public school system, is the weight placed upon disconnected data and procedures that are expected to be acquired at a proficient level of recall by students (similar to the design of computers). And it must be accomplished on a highly structured time schedule that is relatively unforgiving – that is: fall behind and you may never catch up. And the fact that the data and procedures are largely disconnected means that memory is the primary attribute students must rely on if they hope to be academically successful. We must therefore wonder what the real purpose of public education is since it is largely irrational and illogical given the disconnected silos. Curriculum is mandated by the States who do not communicate with the job market. Therefore, State standards should be dissected to determine what they are actually meant to accomplish, besides making colleges'

happy. In addition, colleges do a very poor job of communicating with the job market which is evidenced by the very low confidence companies have in graduates.

It's time citizens take matters in their own hands and embrace home- or micro-schooling.

Recommended Entertaining "Homework" to Close This Essay

To observe an interesting exchange between two people as it relates to the relevance of assessing individuals in public schools, see the 1945 movie *The Bells of St. Mary's* starring Ingrid Bergman (Sister Benedict, school teacher) and Bing Crosby (Father O'Malley, school principal). The scene where Sister Benedict gave failing marks to her 8th grade pupil Patsy based on the principle of standards – which means Patsy would have to repeat 8th grade – and Father O'Malley's opinion that grades aren't as significant as the Sister believes them to be, provides an excellent dialogue of the educational dichotomy. This raises the question the audience must struggle with: Are assessments, grades, and standards important and if so, how are they to be designed and implemented to truly serve individuals – of which, there is an unlimited variety – and society?

In addition, take notice of the real lessons Sister Benedict and Father O'Malley offer Patsy at different points in the movie. These are the most important lessons of all – they are life lessons that will stay with Patsy long after she loses all memory of the data she memorized for tests.

Appendix

Gould on Binet and the Birth of IQ Testing in the U.S.

The following is an outstanding summary of early Progressives' efforts to categorize individuals in order to socially reengineer society based on the German Statist model. Gould (1996) does an outstanding job of laying out how it all unfolded. However, it is obvious that Gould is a Liberal and therefore avoids identifying Progressivism as the source of these evils, which are identified in other sources such as Kliebard (2004).

The following is verbatim from Gould's book.

Pages 179-82
In 1904 Binet was commissioned by the minister of public education to perform a study for a specific, practical purpose: to develop techniques for identifying those children whose lack of success in normal classrooms suggested the need for some form of special education. Binet chose a purely pragmatic course. He decided to bring together a large series of short tasks, related to everyday problems of life…. Learned skills like reading would not be treated explicitly. … Unlike previous tests designed to measure specific and independent "faculties" of mind, Binet's scale was a hodgepodge of diverse activities. He hoped that by mixing together enough tests of different abilities he would be able to abstract a child's general potential with a single score. Binet emphasized the empirical nature of his work with a famous dictum: "One might almost say, 'It matters very little what the tests are so long as they are numerous.'"

Binet published three versions of the scale before his death in 1911. The original 1905 edition simply arranged the tasks in an ascending order of difficulty. The 1908 version established the criterion used in measuring the so-called IQ ever since. Binet decided to assign an age level to each task, defined as the youngest age at which a child of normal intelligence should be able to complete the task successfully. A child began the Binet test with tasks for the youngest age and proceeded in sequence until he could no longer complete the tasks. The age associated with the last tasks he could perform became his "mental age," and his general intellectual level was calculated by subtracting this mental age from his true chronological age. Children whose mental ages were sufficiently behind their chronological ages could then be identified for special educational programs, thus fulfilling Binet's charge from the ministry. In 1912 the German psychologist W. Stern argued that mental age should be divided by chronological age, not subtracted from it, and the intelligence *quotient*, or IQ, was born.

IQ testing has had momentous consequences in our century. In this light, we should investigate Binet's motives, if only to appreciate how the tragedies of misuse might have been avoided if its founder had lived and his concerns been heeded.

…Binet explicitly declined to award any theoretical interpretation to his scale of intelligence, the most extensive and important work he had done in his favorite subject. Why should a great theoretician have acted in such a curious and apparently contradictory way?

Binet did seek "to separate natural intelligence and instruction" in his scale: "It is the intelligence alone that we seek to measure, by disregarding in so far as possible, the degree of instruction which the child possesses…. We give him nothing to read, nothing to write, and submit him to no test in which he might

succeed by means of rote learning" "It is a specially interesting feature of these tests that they permit us, when necessary, to free a beautiful native intelligence from the trammels of the school."

Yet, beyond this obvious desire to remove the superficial effects of clearly acquired knowledge, Binet declined to define and speculate upon the meaning of the score he assigned to each child. Intelligence, Binet proclaimed, is too complex to capture with a single number. This number, later called IQ, is only a rough, empirical guide constructed for a limited, practical purpose:

> The scale, properly speaking, does not permit the measure of the intelligence, because intellectual qualities are not superposable [sic], and therefore cannot be measured as linear surfaces are measured.

Moreover, the number is only an average of many performances, not an entity unto itself. Intelligence, Binet reminds us, is not a single, scalable thing like height. "We feel it necessary to insist on this fact," Binet cautions, "because later, for the sake of simplicity of statement, we will speak of a child of 8 years having the intelligence of a child of 7 or 9 years; these expressions, if accepted arbitrarily, may give place to illusions." Binet was too good a theoretician to fall into the logical error that John Stuart Mill had identified – "to believe that whatever received a name must be an entity or being, having an independent existence of its own."

Binet also had a social motive for his reticence. He greatly feared that his practical device, if reified[48] as an entity, could be perverted and used as an indelible label, rather than as a guide for identifying children who needed help. He worried that schoolmasters with "exaggerated zeal" might use IQ as a convenient excuse: "They seem to reason in the following way: 'Here is an excellent opportunity for getting rid of all the children who trouble us,' and without the true critical spirit, they designate all who are unruly, or disinterested in the school." But he feared even more what has since been called the "self-fulfilling prophesy." A rigid label may set a teacher's attitude and eventually divert a child's behavior into a predicted path:

> It is really too easy to discover signs of backwardness in an individual when one is forewarned. This would be to operate as the graphologists did who, when Dreyfus was believed to be guilty, discovered in his handwriting signs of a traitor or a spy."

Not only did Binet decline to label IQ as inborn intelligence; he also refused to regard it as a general device for ranking all pupils according to mental worth. He devised his scale only for the limited purpose of his commission by the ministry of education: as a practical guide for identifying children whose poor performance indicated a need for special education – those who we would today call learning disabled or mildly retarded. Binet wrote: "We are of the opinion that the most valuable use of our scale will not be its application to the normal pupils, but rather to those of inferior grades of intelligence." As to the causes of poor performance, Binet refused to speculate. His tests, in any case, could not decide:

> Our purpose is to be able to measure the intellectual capacity of a child who is brought to us in order to know whether he is normal or retarded. We should therefore study his condition at the

[48] Make (something abstract) more concrete or real. https://en.oxforddictionaries.com/definition/reify

> time and that only. We have nothing to do either with his past history or with his future; consequently, we shall neglect his etiology, and we shall make no attempt to distinguish between acquired and congenital idiocy…. As to that which concerns his future, we shall exercise the same abstinence; we do not attempt to establish or prepare a prognosis, and we leave unanswered the question of whether this retardation is curable, or even improvable. We shall limit ourselves to ascertaining the truth in regard to his present mental state.

But of one thing Binet was sure: whatever the cause of poor performance in school, the aim of his scale was to identify in order to help and improve, not to label in order to limit. Some children might be innately incapable of normal achievement, but all could improve with special help.

Pages 183-84
How can we help a child if we label him as unable to achieve by biological proclamation?

> If we do nothing, if we don't intervene actively and usefully, he will continue to lose time … and will finally become discouraged. The situation is very serious for him, and since his is not an exceptional case (since children with defective comprehension are legion), we might say that it is a serious question for all of us and for all of society. The child who loses the taste for work in class strongly risks being unable to acquire it after he leaves school.

Binet railed against the motto "stupidity is for a long time" … and upbraided teachers who "are not interested in students who lack intelligence. They have neither sympathy nor respect for them, and their intemperate language leads them to say such things in their presence as 'This is a child who will never amount to anything … he is poorly endowed … he is not intelligent at all.' How often have I heard these imprudent words?" Binet then cites an episode in his own baccalaureate when one examiner told him that he would never have a "true" philosophical spirit: "Never! What a momentous word. Some recent thinkers seem to have given their moral support to these deplorable verdicts by affirming that an individual's intelligence is fixed quantity, a quantity that cannot be increased. We must protest and react against this brutal pessimism; we must try to demonstrate that it is founded upon nothing."

The children identified by Binet's test were to be helped, not indelibly labeled. Binet had definite pedagogical suggestions, and many were implemented. He believed, first of all, that special education must be tailored to the individual needs of disadvantaged children: it must be based on "their character and their aptitudes, and on the necessity for adapting ourselves to their needs and their capacities."[49] Binet recommended small classrooms of fifteen to twenty students, compared with sixty to eighty then common in public schools catering to poor children. In particular, he advocated special methods of education, including a program that he called "mental orthopedics":

> What they should learn first is not the subjects ordinarily taught, however important they may be; they should be given lessons of will, of attention, of discipline; before exercises in grammar, they need to be exercised in mental orthopedics; in a word they must learn how to learn.[50]

---

[49] This is the case for every individual, not just learning disabled students. This is a foundational principle of the Applied Education Foundation.

[50] Learning how to learn is more important than most information that goes by the name of educational material. Most material is spoon-fed to students, which leads to students becoming dependent on institutional structures to

Page 185

The dismantling of Binet's intentions in America

In summary, Binet insisted upon three cardinal principles for using his tests. All his caveats were later disregarded, and his intentions overturned, by the American hereditarians who translated his scale into written form as a routine device for testing all children.

1. The scores are a practical device; they do not buttress any theory of intellect. They do not define anything innate or permanent. We may not designate what they measure as "intelligence" or any other reified entity.
2. The scale is a rough, empirical guide for identifying mildly retarded and learning-disabled children who need special help. It is not a device for ranking normal children.
3. Whatever the cause of difficulty in children identified for help, emphasis shall be placed upon improvement through special training. Low scores shall not be used to mark children as innately incapable.

If Binet's principles had been followed, and his tests consistently used as he intended, we would have been spared a major misuse of science in our century. Ironically, many American school boards have come full cycle, and now use IQ tests only as Binet originally recommended: as instruments for assessing children with specific learning problems. …

Page 187

American [Progressive] psychologists perverted Binet's intention…. They reified Binet's scores, and took them as measures of an entity called intelligence. … They believed that inherited IQ scores marked people … for an inevitable station in life. …

This chapter analyzes the major works of the three pioneers of hereditarianism in America: H. H. Goddard, who brought Binet's scale to America and reified its scores as innate intelligence; L. M. Terman, who developed the Stanford-Binet scale, and dreamed of a rational society that would allocate professions by IQ scores; and R. M. Yerkes, who persuaded the army to test 1.75 million men in World War I….

Page 189

Goddard was the first popularizer of the Binet scale in America. He translated Binet's articles into English, applied his tests, and agitated for their general use. He agreed with Binet that the tests worked best in identifying people just below the normal range – Goddard's newly christened morons. But the resemblance between Binet and Goddard ends there. Binet refused to define his scores as "intelligence," and wished to identify in order to help. Goddard regarded the scores as measures of a single, innate entity. He wished to identify in order to recognize limits, segregate, and curtail breeding to prevent further deterioration of an endangered American stock, threatened by immigration from without and by prolific reproduction of its feeble–minded within.

Pages 190-92

---

learn rather than becoming independent learners. This leads to a populace dependent upon government rather than upon their own abilities.

Goddard extended the range of social phenomena caused by differences in innate intelligence until it encompassed almost everything that concerns us about human behavior. Beginning with morons, and working up the scale, he attributed most undesirable behavior to inherited mental deficiency of the offenders. Their problems are caused not only by stupidity per se, but by the link between deficient intelligence and immorality.

At the next level of the merely dull, we find the toiling masses, doing what comes naturally. "The people who are doing the drudgery," Goddard writes, "are, as a rule, in their proper places."

> We must next learn that there are great groups of men, laborers, who are but little above the child, who must be told what to do and shown how to do it; and who, if we would avoid disaster, must not be put into positions where they will have to act upon their own initiative or their own judgment…. There are only a few leaders, most must be followers.[51]
>
> … How can there be such a thing as social equality with this wide range of mental capacity? [52]

"Democracy," Goddard argued, "means that the people rule by selecting the wisest, most intelligent and most human to tell them what to do to be happy. Thus Democracy is a method for arriving at a truly benevolent aristocracy."[53]

But if intelligence forms a single and unbroken scale, how can we solve the social problems that beset us? For at one level, low intelligence generates sociopaths, while at the next grade, industrial society needs docile and dull workers to run its machinery and accept low recompense. How can we convert the unbroken scale into two categories at this crucial point, and still maintain the idea that intelligence is a single, inherited entity? We can now understand why Goddard lavished so much attention upon the moron. The moron threatens racial health because he ranks highest among the undesirable and might, if not identified, be allowed to flourish and propagate. We all recognize the idiot and imbecile and know what must be done; the scale must be broken just above the level of the moron.

> The idiot is not our greatest problem. He is indeed loathsome…. Nevertheless, he lives his life and is done. He does not continue the race with a line of children like himself…. It is the moron type that makes for us our great problem.

… Eugenicists … allowed them to assert that all undesirable traits might be traced to single genes and eliminated with proper strictures upon breeding. … We must not be misled by how silly such ideas seem today; they represented orthodox genetics for a brief time, and had a major social impact in America.[54]

---

[51] Here we see the origins of the educational establishment's view that higher education's purpose is to mold leaders of the future. The rest of the citizenry is to be relegated to be part of the "toiling masses." This helps explain why "college for all" is really "college for the few who can make it" and the rest will be laborers led by those who "made it." The costs incurred by those who failed are justified in order to discover and then mold the "leaders."

[52] When one group optimizes the social system based on its own talents – such as academics have, for example – there can be no equality. After all, equality degrades the elevated status of the "elite" group.

[53] This is a good summary of Progressive-Liberalism rooted in Statism. However, democracy does not come close to picking "the wisest, most intelligent and most human" people to manage the affairs of State. It tends to pick narcissists that are masters of sophistry as was seen in ancient Greece. An important distinction to make: The U.S. is not a democracy, as Greece was, it is a republic, as Rome was before the Republic was lost to empire when Caesar took control.

Page 193
Goddard claimed that he had been compelled to make this unlikely conclusion by the press of evidence, not by any prior hope or prejudice.

Any theories or hypotheses that have been presented have been merely those that were suggested by the data themselves, and have been worked out in an effort to understand what the data seem to comprise. Some of the conclusions are as surprising to the writer and as difficult for him to accept as they are likely to be to many readers.[55]

If mental deficiency is the effect of a single gene, the path to its eventual elimination lies evidently before us: do not allow such people to bear children:

> If both parents are feeble-minded all the children will be feeble-minded. It is obvious that such matings should not be allowed. It is perfectly clear that no feeble-minded person should ever be allowed to marry or to become a parent. It is obvious that if this rule is to be carried out the intelligent part of society must enforce it.

Page 194

> They are not only lacking in control but they are lacking often in the perception of moral qualities; if they are not allowed to marry they are nevertheless not hindered from becoming parents. So that if we are absolutely to prevent a feeble-minded person from becoming a parent, something must be done other than merely prohibiting the marrying. To this end there are two proposals: the first is colonization, the second is sterilization.[56]

Page 204
Lewis M. Terman and the mass marketing of innate IQ
Walter Lippmann, in the course of a debate with Lewis Terman [said]:

> Without offering any data on all that occurs between conception and the age of kindergarten, they announce on the basis of what they have got out of a few thousand questionnaires that they are measuring the hereditary mental endowment of human beings. Obviously, this is not a conclusion obtained by research. It is a conclusion planted by the will to believe. It is, I think, for the most part unconsciously planted…. If the impression takes root that these tests really measure intelligence, that they constitute a sort of last judgment on the child's capacity, that they reveal "scientifically" his predestined ability, then it would be a thousand times better if all the intelligence testers and all their questionnaires were sunk without warning in the Sargasso Sea.

Pages 205-06

---

[54] Here Gould fails to tell the whole story. This was not an "American" effort; it was a Progressive effort.

[55] This is a perfect example of how far off science can be. American's adoration of science needs to be tempered by equal respect for intuition. Logic is a tool, not the absolute answer to discovering truths. The great minds of math and science have been fully aware of this reality.

[56] Here are the origins of the Nazis' view of a superior race. In fact Hitler gave U.S. Progressives full credit for his perspective on eugenics, which eventually evolved into genocide.

Goddard introduced Binet's scale to America, but Terman was the primary architect of its popularity. … Terman's first revision of 1916 extended the scale to "superior adults"…. Terman, by then a professor at Stanford University, gave his revision a name that has become part of our century's vocabulary – the Stanford-Binet, the standard for virtually all "IQ" tests that followed.

… When expectations are society's norms, then do the tests measure some abstract property of reasoning, or familiarity with conventional behavior?

Pages 206-07
But Terman's major influence did not reside in his sharpening or extension of the Binet scale. Binet's tasks had to be administered by a trained tester working with one child at a time. They could not be used as instruments for general ranking. But Terman wished to test everybody, for he hoped to establish a gradation of innate ability that could sort all children into their proper stations in life:

> What pupils shall be tested? The answer is, all. If only selected children are tested, many of the cases most in need of adjustment will be over-looked. The purpose of the tests is to tell us what we do not already know, and it would be a mistake to test only those pupils who are recognized as obviously below or above average. Some of the biggest surprises are encountered in testing those who have been looked upon as close to average in ability. Universal testing is fully warranted.

Pages 211-212
Feeble-minded people are doubly burdened by their unfortunate inheritance, for lack of intelligence, debilitating enough in itself, leads to immorality.

> Not all criminals are feeble-minded, but all feeble-minded persons are at least potential criminals. … The feeble-minded, in the sense of social incompetents, are by definition a burden rather than an asset, not only economically but still more because of their tendencies to become delinquent or criminal…. The only effective way to deal with the hopelessly feeble-minded is by permanent custodial care. The obligations of the public school rest rather with the large and more hopeful group of children who are merely inferior.

Terman virtually closed professions of prestige and monetary reward to people with IQ below 100, and argued that "substantial success" probably required an IQ above 115 or 120.

… IQ of 75 or below should be the realm of unskilled labor, 75 to 85 "preeminently the range for semi-skilled labor." … Proper vocational training and placement is essential for those "of the 70 to 85 class." Without it, they tend to leave school "and drift easily into the ranks of the anti-social or join the army of Bolshevik discontents."

Pages 220-21
Terman moved easily from individuals, to social classes, to races. Distressed by the frequency of IQ scores between 70 and 80, he lamented:

> Among laboring men and servant girls there are thousands like them. … The tests have told the truth. These boys are ineducable beyond the merest rudiments of training. No amount of school

instruction will ever make them intelligent voters or capable citizens…. Their dullness seems to be racial, or at least inherent in the family stocks from which they came.[57]

[57] Here we hear a Progressive speaking in highly bigoted terms, which was the norm until the Progressives learned of the Nazis' genocide efforts at which point they abandoned eugenics.

References

AACC Staff, *Diagnosing College Readiness as Early as Middle School*, April 24, 2015.

AAC&U, *On Solid Ground*, Value Report 2017, Association of American Colleges and Universities.

Abdul-Alim, Jamaal, *Expert: Higher Ed Needs to Embrace Assessments of Student Learning*, Diverse Issues in Higher Education, Mar. 31, 2016.

Ashford, Ellie, *Cutting Back on Remediation Yields Success*, Community College Daily, April 26, 2017.

Belfield, Clive, Jenkins, Davis, and Lahr, Hana, *Is Corequisite Remediation Cost-Effective? Early Findings From Tennessee*, CCRC Research Brief No. 62, Community College Research Center, Teachers College, Columbia Univ., April 2016.

Bethke (2016) *New Type of Remediation is Making Waves in Higher Ed*, eCampus News, Feb. 4, 2016.

Bickford, Ian, *The College Prep That's Worse Than Neutral? High School: How "Early College" Reimagines the Path to Higher Ed*, The Hechinger Report, Dec. 27, 2017.

Bucholz, Robert, *A History of England From the Tudors to the Stuarts*, Lectures 21 – 23, The Great Courses, The Teaching Company, 2003.

Burke, Michael, *Letter Grades on Way Out? Why Some University of California Departments May Use Alternatives*, EdSource, April 26, 2022.

Carnegie, Andrew, *The Empire of Business*, Doubleday, Page & Co., 1902.

Carnoy, M., Elmore, R., and Siskin, L., editors, *The New Accountability: High Schools and High-Stakes Testing*, Routledge Falmer, 2003.

Cohen, G., Dweck, C., and Walton, G., *Academic Tenacity: Mindsets and Skills that Promote Long-Term Learning*, Bill & Melinda Gates Foundation.

Committee on the Foundations of Assessment: Pellegrino, Chudowsky, and Glaser, editors, Board on Testing and Assessment, Center for Education, National Research Council, *Knowing What Students Know: The Science and Design of Educational Assessment*, National Academy of Sciences, National Academy Press, 2001.

Conley, David, *A New Era For Educational Assessment*, Jobs for the Future, Oct. 2014.

Devaney, Laura, *Illinois Community Colleges are Reforming Developmental Education*, eCampus News, July 27, 2016.

Dierking, Phil, *New System to Measure 'Value' of College Education*, Learning English VOA News, April 08, 2017.

Edublox, IQ Test: Where Does it Come From and What Does it Measure? https://edubloxtutor.com/history-iq-test/.

Edley, C., Jr. and Koenig, J., editors. National Academies of Sciences, Engineering, and Medicine. (2017). *Evaluation of the Achievement Levels for Mathematics and Reading on the National Assessment of Educational Progress*. Wash. D.C.: The National Academies Press. doi: 10.17226/23409.

Fordham Institute, Petrilli, M. and Finn, C., *College Preparedness Over the Years, According to NAEP*, April 8, 2015.

Gardner, Howard, *Frames of Mind*, Basic Books, 1983.

Goleman, Daniel, *Emotional Intelligence: Why It Can Matter More Than IQ*, Bantam Books, 1995.

Goleman, Daniel, *Social Intelligence*, Bantam Books, 2006.

Goodman, M. J., Sands, A. M., and Coley, R. J., *America's Skills Challenge: Millennials and the Future*, Educational Testing Service, 2015.

Gottipati, Swapna and Shankararaman, Venky. Competency analytics tool: Analyzing curriculum using course competencies. (2018). Education and Information Technologies. 23, (1), 41-60. Research Collection School of Information Systems.

Gould, Stephen, *The Mismeasure of Man*, W.W. Norton & Co., 1996.

Guskey, Thomas, *On Your Mark: Challenging the Conventions of Grading and Reporting*, Solution Tree Press, 2015.

Hancock, LynNell, *Why Are Finland's Schools Successful?*, Smithsonian Magazine, September 2011.

Hechinger Report, *Graduates of 4-Year Universities Flock to Community Colleges for Job Skills*, U.S. News & World Report, Oct. 28, 2015.

Herrnstein, Richard and Murray, Charles, *The Bell Curve*, Free Press Paperbacks, 1996.

Hoffman, Banesh, *The Tyranny of Testing*, Dover, 1964/2003.

How People Learn II: Learners, Contexts, and Cultures, The National Academies of Sciences, Engineering, and Medicine, 2018.

Hunt, McVicker, *Intelligence and Experience*, The Ronald Press Company, 1961.

Janio, J. and Gaff, D., "From Apprenticeship to Competency: What Anthropology Can Teach Us About Learning," *The Evolllution*, May 28, 2025.

Janio, J., "Completion is Not Competence: Why Accountability Requires a Definition of Learning," *The Evolllution*, Mar. 12, 2026.

Jaschik, Scott, *ACT Scores Drop as More Take Test*, Inside Higher Ed, August 24, 2016.

Johnson, Derek, *SAT, ACT Scores Alone Not Enough to Predict First-Year College Success, Says New Study*, Education News, July 15, 2016.

Kallendorf, *Humanist Educational Treatises*, The I Tatti Renaissance Library, Harvard Univ. Press, 2002.

Kaufman, Alan, *IQ Testing 101*, Springer Publishing Company, 2009.

Kiener, A., Kragthorpe, H., Randall, E., Banovetz, N., Fan, A., Varro, V., Smith, D., Chang, J., Davis, B., and Evans, R., *ACT Debate Misses the Point: Let's Get Serious About College Access*, Minncan, Aug. 11, 2015.

Kliebard, Herbert, *The Struggle for the American Curriculum: 1893–1958*, RoutledgeFalmer, NY, 2004.

Koretz, D., Linn, R. Dunbar, S. & Shepard, L., *The Effects of High-Stakes Testing on Achievement: Preliminary Findings About Generalization Across Tests*, Harvard, Sept. 2, 2014.

Lave, Jean, *Cognition in Practice*, Cambridge Univ. Press, 1988.

Lemann, Nicholas, *The Great Sorting*, The Atlantic, Sept. 1995.

Lemann, Nicholas, *The Structure of Success in America*, The Atlantic, Sept. 1995.

Logue, A. W., Rose, M., and Douglas, D., *Should Students Assessed as Needing Remedial Mathematics Take College-Level Quantitative Courses Instead?*, American Educational Research Association, *Evaluation and Policy Analysis*, Vol. XX, No. X, pp. 1-21, 2016.

Louis Blanc, *Letters on England*, Vol. II, Sampson Low, Son, and Marston, Milton House, Ludgate Hill, London, 1866.

Lubbers, Teresa, *State's Program to Prepare Students is Getting Results – and Notice*, The Journal Gazette, Feb. 10, 2016.

Lubischer, Jane, "Our Transcripts Are Academic Rap Sheets – and We Can Do Better," *Inside Higher Ed*, Oct. 3, 2023.

Lynch, Matthew, *Three Important Critiques of Standardized Assessments*, Education Week, June 28, 2016.

Mangan, Katherine, *Making Transcripts More Than 'A Record of Everything the Student Has Forgotten*, The Chronicle of Higher Education, July 15, 2015.

Marcus, Jon, *Competency-Based Education, Put to the Test*, Education Next, Vol. 17, No. 4, Fall 2017.

Martin, Robert and Gillen, Andrew, Can We Measure the Value of College Teaching?, Minding the Campus, Feb. 7, 2011.

Massachusetts Business Alliance for Education in conjunction with Center for Assessment, *Educating Students for Success: A Comparison of the MCAS and PARCC Assessments as Indicators of College- and Career-Readiness*, Feb. 2015.

Mestre, Jose, Ed., *Transfer of Learning: From a Modern Multidisciplinary Perspective*, Information Age Publishing, 2005.

Meyers, Jason, *Validating Competencies and Skills Gained Through Rigorous Assessment*, Western Governors University, The Evolllution, Oct. 31, 2019.

Mintz, Steven, *Rethinking Grades*, Inside Higher Ed, Mar. 2, 2016.

Mintz, Steven, *Reimagining the College Transcript*, Inside Higher Ed, Jan. 12, 2017.

Mortrude, Judy, *GED Scoring Change Provides Opportunity for Career Pathway Alignment*, Clasp, Jan. 29, 2016.

National Education Association, *Cardinal Principals of Secondary Education: A Report of the Commission on the Reorganization of Secondary Education*, 1918.

National Research Council, *How People Learn: Brain, Mind, Experience, and School*, 2000.

National Research Council. 2001. *Knowing What Students Know: The Science and Design of Educational Assessment*. Washington, DC: The National Academies Press.

National Research Council. 2012. *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: The National Academies Press. https://doi.org/10.17226/13398.

National Research Council, *How People Learn II: Learners, Contexts, and Cultures*, The National Academies Press, 2018.

Nelson, Christopher, *Assessing Assessment*, Inside Higher Ed, Nov. 24, 2014.

NWEA, *Make Assessment Work for All Students: Multiple Measures Matter*, Gallup, 2016.

Osborn, Eliana, *First Semester GPA Predicts College Success Better Than Standardized Test Scores*, Education News, March 14, 2016.

Palmer, Iris, *How to Fix Remediation at Scale*, New America, March 2016.

Pena, Adolfo, *The Dreyfus Model of Clinical Problem-Solving Skills Acquisition: A Critical Perspective*, VAQS, Birmingham VA Medical Center, Univ. of Alabama at Birmingham, Medical Education Online 2010, 15: 4846 - DOI: 10.3402/meo.v15i0.4846

Peterson, Jordan, *The Neuroscience of Intelligence: Dr. Richard Haier* interview, https://www.youtube.com/watch?v=PY4sShDt9to, 2017.

Preville, Philip, *The Professor's Guide to Using Bloom's Taxonomy*, tophat.com, year?

Quick, Robert H., *Essays on Educational Reformers*, D. Appleton & Co., 1894.

Rutter, Michael, *Performance-Based Assessment*, Inside Higher Ed, April 29, 2015.

Schaeffer, Bob, *Test-Optional Admissions Growth Surge: New Survey Shows Record Number of Colleges and Universities Dropped ACT/SAT Exam Score Requirements in the Past Year: List Now Includes 850 Bachelor-Degree Granting* Schools, Fair Test, April 29, 2015.

Schwartz, D. L., Bransford, J. D., and Sears, D., *Transfer of Learning: From a Modern Multidisciplinary Perspective*, Ed. Jose Mestre, Chapter 1, Efficiency and Innovation in Transfer, Information Age Publishing, 2005.

Selingo, Jeffrey, *Why Are So Many College Students Failing to Gain Job Skills Before Graduation?*, The Washington Post, Jan. 26, 2015.

Smith, Ashley, *College Students Placed in Remedial Algebra Have Better Outcomes in College Stats Classes*, Inside Higher Ed, June 23, 2016.

Sparks, Sarah, *Should Schools Test the "Career" Half of "College and Career"?*, Education Week, May 24, 2017a.

Sparks, Sarah, *What 150 Years of Education Statistics Say About Schools Today*, Education Week, Nov. 16, 2017b.

Spencer, Herbert, *Education: Intellectual, Moral and Physical*, A. L. Burt Co., 1860.

Stanovich, Keith, *What Intelligence Tests Miss: The Psychology of Rational Thought*, Yale Univ. Press, 2009.

Stansbury, Meris, *Replacing the SATs in admissions–what comes next?*, eCampus News, March 16, 2016.

Strong, Sarah, *Making Math About More Than Numbers*, Education Week, Nov. 15, 2017.

Travis, Scott, *Colleges See Some Success Since Slashing Remedial Classes*, Sun Sentinel, Oct. 29, 2016.

U.S. Army, *Army Talent Assessment Strategy*, Dept. of the Army, Jan. 31, 2024.

Walsh, Kelly, *CBE + Stackable Credentials + Blockchain Transcripts = Higher Ed Flexibility*, EmergingEdTech.com, 2017.

Wieman, Carl, *A Better Way to Evaluate Undergraduate Teaching*, Change: The Magazine of Higher Learning, Jan.-Feb. 2015.

Willingham, Daniel, *Cognition: The Thinking Animal*, Pearson Education, 2007.