

Psychometric Superiority? Check the Facts—Again

Mindy L. Kornhaber
The Pennsylvania State University
Kornhaber@psu.edu
August 15, 2004

Daniel Willingham's critique of Howard Gardner's theory of multiple intelligences (Summer 2004) is yet another attempt by a psychometric supremacist to quash other views of intelligence. Willingham claims that Gardner's theory "is almost certainly incorrect," and that the current psychometric view almost certainly is, since it fits the data. Let us first examine that view and then ponder an issue that mystifies Willingham: Namely, why do so many educators, despite the "science" of psychometrics, instead embrace Gardner's theory?

One of the reasons that psychometric diehards continue to get undue attention is that the public has very little idea about how psychometric models are built. All such models are based on the spectacularly flawed assumption that the nature of intelligence can best be understood through analyses of how people think in highly atypical situations. It is a fact, as Willingham has stated, that data from some 130,000 individuals enabled John Carroll to produce a three-tiered hierarchical model of intellect—with *g* (general intelligence) atop it all. But here are three more facts: These 130,000 individuals were largely sitting in isolated rows frenetically wielding number 2 pencils over small booklets of peculiar puzzles and bubble-in answer sheets. Nobody invests any sizable part of his or her intellectual life in such activity (or its contemporary computer-equivalents). Not a single intellectual product of any merit has ever resulted from that setting. Thus, a model of intelligence built on that activity, no matter how large the sample size and how rigorously statistics contributed to its assembly, is, at best, extremely limited. The model may fit the data just beautifully, but the data aren't drawn from the reality of how human beings actually use their minds.

A dull roar rises from the diehard crowd in response to these sorts of arguments. They counter that intelligence test scores are associated with important real-world events, like job status and income. They also argue that more recent neurological studies shore up their approach and give it a biological foundation. To grasp the problems with such arguments, consider the following scenario: A model of nutrition based on anorexics' blood samples would likely be correlated with such important real-world health issues as low rates of diabetes and cancer. It's also likely that biologically-based, causal explanations could be found to support such a statistical relationship (e.g., perhaps low demands on the pancreas or reduction of free radicals). Yet, such correlations or even causal explanations would still only leave us with a model of starvation and not nutritional health. In essence, a model of intelligence driven by test puzzles and bubble sheets is about as good as a model of nutrition based on anorexics' blood samples. It reflects the intellectual equivalent of starvation.

Gardner's theory provides a much-needed corrective to the shortcomings of traditional psychometric approaches. Instead of probing the bases of bubble-sheet

results, Gardner sought to illuminate the mental abilities underlying the range of actual human accomplishments that are found across cultures. How is it that there are not only capable mathematicians and writers, but also skillful historians, farmers, engineers, actors, teachers, lawyers, dancers, hunters, politicians, and, yes, even comedians?

Gardner tackled this question by drawing on evidence that psychometricians have ignored. Indeed, it is the use of a wide array of evidence that distinguishes Gardner from many other investigators who have argued in favor of multiple intellectual capabilities. Drawing on research in cognitive developmental psychology, Gardner considered whether or not abilities have separate developmental trajectories from infancy onward. (One question regarding the possibility of comedic intelligence: Do cartoonists, comedy writers, and slapstick artists develop at a similar rate or do they develop in ways that parallel painters, poets, and dancers?) Drawing on the available work in neuropsychology, Gardner looked for patterns of intellectual exceptionality and deficits among prodigies, those with brain injuries, and other special populations. (Neuropsychological evidence could be applied to the possibility of comedic intelligence: If a comedy writer suffered a stroke and could no longer write comedy, could he produce or appreciate comedy in other forms? Gardner's own work in clinical neurology revealed cases of brain-injured people who had lost their ability to appreciate humor in one modality while maintaining it in others. Oliver Sacks reported similar findings. Hence, Willingham's argument about humor intelligence gets the hook.) In addition, Gardner considered whether various mental abilities warranted the invention of an attendant symbol system. He looked for evidence of distinct abilities from evolutionary biology, experimental psychology, and in findings from psychometric psychology. He also went beyond psychometric approaches by identifying computational systems that operate on information that enters via the several senses.

Statistical techniques for analyzing and synthesizing this diversity of evidence do not exist. Therefore, Gardner could not number crunch some data, get a $p=.05$ and exclaim that there was a statistically significant "finding," and then crunch other data, get a $p=.06$ and note, alas, that there was not. (The statistically over-committed see this distinction as a "convention" rather than "arbitrary." That they do is but one of many examples illustrating how human judgment actually pervades the work of those purporting to have the "objective" truth about intelligence.) Given that Gardner's evidence was not susceptible to meaningful statistical analyses, he made it plain that his analysis relied on human judgment. He laid out his evidence, criteria, and reasoning (FYI, so did Darwin). In this fashion, Gardner ultimately arrived at eight different intelligences.

It is perfectly fine for Willingham to consider the possible existence of other intelligences. In fact, since the theory's inception in 1983, Gardner has asserted that if new evidence arises and his criteria can be met, the set of intelligences may be modified. Gardner's theory has since been supported by newer work in evolutionary biology and psychology, through which separate modules, resembling the candidate intelligences, have been identified. At the same time, studies in the late 1990s that link musical and spatial intelligence may be seen as counterevidence for the relative autonomy of those

two intelligences. Ironically, the study by Carroll that Willingham dwells on has humbled the once-mighty stance of general intelligence, and thus also serves as an additional support for Gardner's theory. Carroll's study is an odd cudgel for Willingham to wield against Gardner for yet another reason: Carroll himself regarded the theory of multiple intelligences as useful (Check Carroll's *Human Cognitive Abilities*, pp. 641-642).

A number of Willingham's arguments are conspicuously ahistorical and factually incorrect. As no doubt a few other readers recognized, it is misleading to fault Gardner for not drawing on Carroll's study, which was published only a decade *after* Gardner produced his theory. Chronology fails Willingham again when he indicates that Carroll's work was the nail in *g*'s coffin, since praises to *g* continued thereafter. (e.g., Herrnstein and Murray's 1994 book, *The Bell Curve* and its many sequelae.) Willingham is also slippery in using Thurstone's earlier psychometric work to make Gardner's theory appear tepid. First, Thurstone relied on test data, rather than diverse sources of evidence, to arrive at his set of seven primary mental abilities. Second, Thurstone's set of abilities was widely dismissed by mainstream psychometricians, and ultimately Thurstone himself believed that it did not topple *g*. Thus, when the chronology and facts are correctly laid out, those psychometricians who have lately taken to skittering away from *g* only gird Gardner's original challenge to it.

Debates about the nature of intelligence are not "sound and fury, signifying nothing," either for educators, as Willingham claims, or, I would add, for anyone else. Portraying them in that light masks a crucial, underlying issue: Who has the authority to decide what intelligence is and therefore the privilege to deem some people intelligent and others not? Historically, those on top of the pile have preferred conceptions of intelligence that shore up their own power base. Thus, the emperors of China valued and rewarded men whose minds most eloquently spat back Confucian notions of hierarchy. Francis Galton, an upper-class Englishman, member of the Darwin family, and arguably the first scientific investigator of intelligence, found that upper-class Englishmen from eminent families were inherently best and brightest. The Harvard, Princeton, and Stanford men who devised mass-administered intelligence tests for World War I army recruits asserted that White Anglo-Saxon Protestants with northern European backgrounds outranked Whites with families originating elsewhere in Europe, Jews, and Blacks. Psychometric supremacists argue that their methods entitle them to the privilege of deeming some smart and some not. Not surprisingly, they value those who do well on tests that they themselves design or study. Those with accomplishments or capabilities that don't shake out via the supremacists' puzzles and procedures must be something *other than* intelligent. Maybe they're merely "talented." Maybe they're even stupid. But here's the rub: Plenty of people who are far from stupid take issue with psychometric views. It must be maddening to the psychometric supremacists, given their refined methods and mighty minds, to find that many talented—and even highly intelligent—people are no longer thoroughly in their thrall.

This brings us to Willingham's mystery: Why is it that the psychometrically overcommitted have so often rejected Gardner's theory while educators have embraced

it? Willingham gives the following explanation: "educators got so confused by Gardner's theory." He claims that they simply fell for a bait and switch of "intelligences" for "talents." Aside from insulting educators' intelligence, this explanation merely mirrors earlier arguments by some other psychometric diehards. For someone so committed to data and its analysis, Willingham's explanation is markedly lacking in both.

Moving from groundless hunch to reason and data does much to clarify Willingham's mystery. Let's consider the possibility that educators were lured to the theory by just a simple name change, from talents to intelligences. Gardner himself has speculated that his terminology may initially attract educators' notice. However, would educators stick with the theory if a name change were all there was to it? Given the faddishness of educational innovation, this is implausible. Instead, multiple intelligences theory would have gone the way of any number of other educational ideas that had a spectacular launch, only to crash and burn not long thereafter. Thus, some other things must help to explain educators' sustained interest.

One of these things is that psychometric models are of scant utility to teachers in ordinary classrooms. The staying power that such models have had in education reflects bureaucratic, rather than pedagogical, needs. Psychometric instruments are typically designed to report that most students are about average and only a very few perform far above or below that level. This design conveniently offers school- and district administrators a quantitative, efficient, and "objective" tool for allocating various educational resources (e.g., special ed, gifted ed) to various students. Alas, it offers the bulk of classroom teachers little guidance beyond "go slower," "go faster"; or perhaps, "change your expectations."

The theory of multiple intelligences also initially offered little that was directly useful to teachers. The theory's popularity among teachers therefore serves as a window onto the larger question of why teachers employ particular ideas and reject others. Having explored educators' yen for the theory for several years, it is clear to me that Willingham's explanation markedly undersells teachers' thinking. Interview data that Mara Krechevsky and I collected from nearly 100 teachers in 11 public school districts in disparate parts of the United States clarify why teachers gravitate to Gardner's theory: It complements educators' existing philosophies and beliefs (e.g., that education should address "the whole child"). It validates what educators already know and do (e.g., students learn in a variety of ways). The theory provides a vocabulary for considering, communicating about, and serving learners with diverse intellectual profiles. The theory fits at least some of their current practice (e.g., efforts to employ interdisciplinary curriculum or hands-on science). It enables teachers to reflect systematically on their practice and thereby enhance it. In short, classroom teachers employ Gardner's theory because it makes sense to them, and it's useful. These reasons, rather than simple mass confusion among teachers, explain the theory's influence over the course of 20 years.

Given that Willingham thinks better practice is associated with more accurate descriptions of the mind, it is exceedingly odd that he offers not a single example of good practice stemming from psychometric views. It's less surprising that he fails to note that

psychometric ideas have been plagued by notoriously bad practices, some directly endorsed by psychometric theorists. For example, psychometric data from intelligence tests were used to restrict US borders to Eastern European immigrants, thereby dooming many to the Holocaust. Psychometric data were also used in the US to identify individuals to undergo forced sterilization procedures, thereby preventing them from possibly reproducing similarly low-scoring children. (Check the facts: *Buck v. Bell* 274 US 200 (1927)). While individually administered psychometric assessments may provide useful data about particular students, the use of group administered psychometric tests, especially intelligence tests, has, in the aggregate, yielded more harm than good. If psychometric views about intelligence are correct and widely agreed upon, as Willingham claims, then Willingham's notion that only a good theory of intelligence can lead to good practice is far off the mark. Conversely, if Willingham's view of theory into practice is correct, then problems must inhere with the psychometric view that he advocates.

Reality is even better than this kind of logic game: Educational theorists and policymakers widely acknowledge that translating theory into practice is rarely straightforward. This is true of psychometric theory. It is also true of Gardner's theory. Unlike many other educational ideas, the theory of multiple intelligences was largely a bottom-up phenomenon, without an umbrella agency to monitor and guide its use in practice. Gardner, who had not anticipated the wide use of his theory in schools, initially encouraged broad experimentation with it. Until quite recently, many who promoted the theory's adoption drew almost exclusively on their own idiosyncratic ideas or on practices enacted in their own classroom or school. Hence, brilliant and vapid uses abounded.

But given that all theories have problems being translated into practice, a more pertinent concern should be this: What is the potential of a given theory to promote good practice? Despite all the methodological wizardry offered by the supremacists and all the problems encountered in translating Gardner's theory into practice, I would argue that the latter nevertheless holds much greater potential. This is not just because of the outright horrors or ongoing problems associated with psychometrically-influenced practice. It is not just because psychometric approaches are employed largely for bureaucratic rather than pedagogical purposes, while Gardner's theory encourages teachers to reflect on and grow their own practice. This assertion also rests on evidence from the study of 41 schools that I conducted with Edward Fierros and Shirley Veenema, which Willingham alluded to. That study was not directly aimed at establishing the theory's "effectiveness." It asked a different question, namely: "What are the practices in place in schools that use multiple intelligences theory and associate it with benefits for students?" The study revealed that schools using the theory were notable for practices that are widely seen as beneficial. For example, the schools commonly had a culture of hard work, respect, and caring; a faculty that collaborated and learned from each other; classrooms that engaged students through constrained but meaningful choices, and a sharp focus on enabling students to produce high-quality work. The theory, according to teachers and principals, helped to support these practices. After a full century of dominance, it remains highly

questionable whether psychometric theories, which are neither the whole story of human intelligence nor terribly useful to classroom teachers, can do anything similar.