

Appendix: Extended Examples of Thirteen Especially Influential Logical Fallacies About Intelligence Testing Note: The bolded text in brackets annotates the quotations.			
Example	Test design fallacy # 1	Yardstick mirrors construct	Portraying the superficial appearance of a test (Entry 8) as if it mimicked the inner essence of the phenomenon it measures (Entry 5).
i	Fischer et al. (1996)	Context: Authors are arguing that the Armed Forces Qualification Test (AFQT) does not measure IQ or “intelligence broadly understood” (p. 43) but only learning in school.	<p>Quotation: Psychometricians did not identify <i>g</i>, the general factor for intelligence, by observing people having intelligently; they derived it [the latent construct] from statistical analyses of test questions, from the tendency of people who answer one question accurately to answer others accurately. It is a concept built from the test upward. In chapter 2, we looked at a few questions from the AFQT itself [concrete aspects of the yardstick]. They clearly tested an examinee’s command of school curricula. Here are a few more examples:</p> <p>Two partners, X and Y, agree to divide their profits in the ratio of their investments. If X invested \$3,000 and Y invested \$8,000, what will be Y’s share of a \$22,000 profit?</p> <p>....As before, we see that the AFQT questions are manifestly about [superficially look like] <i>school</i> tasks (pp. 56-57). ...Our critique here rests on questioning the AFQT’s content validity (see chapter 2) as a test of <i>g</i> [a construct] by simply reading the test [gazing at the yardstick] (p. 58)....Statistical evidence supports reading the AFQT as essentially a test of mastering school curricula [yardstick measures only what its superficial appearances suggest] (p.59)....On face value, these questions do not measure test takers’ intelligence, their “deeper capability...for ‘catching on.’” Mostly they measure test taker’s exposure to curricula in demanding math and English classes. They remind us of pop quizzes in high school (p. 42).</p>
ii	Flynn (2007)	Context: Author is proposing a skills-based definition of intelligence that is	Quotation: As for WISC subtests, Similarities, Block Design, Object Assembly, Picture Arrangement, and Picture Completion all measure mental acuity to some degree. Information [yardstick] and Vocabulary [yardstick] measure what they say [yardstick = construct]. Arithmetic

		“narrow enough to offer good advice to those who want to make intelligence measurable and specific” (p. 55).	measures learning what schools teach as mathematics. Comprehension measures understanding the mechanics of everyday life. Coding and Symbol Search measure processing speed. Forward Digit Span isolates memory from the other components of intelligence broad. My classification of subtests differs from that offered in the WISC manuals...Theirs is based on factor analysis [i.e., identifying latent constructs], mine on matching test content with functional mental processes (p. 55).
iii	Sternberg, Wagner, Williams, & Horvath (1995)	Context: Authors are arguing that different item formats (“academic” vs. “practical”) necessarily require different intelligences. They claim IQ tests use only the former and thus can measure only an “academic intelligence” (g).	Quotation: Neisser (1976) was one of the first psychologists to press the distinction between academic and practical intelligence [proposed constructs of academic intelligence and practical intelligence]. Neisser described academic intelligence tasks (common in the classroom and on intelligence tests) as (a) formulated by others, (b) often of little or no intrinsic interest, (c) having all needed information available from the beginning, and (d) disembedded from an individual’s ordinary experience [yardstick for “academic intelligence”]. In addition, one should consider that these tasks (e) usually are well defined, (f) have but one correct answer, and (g) often have just one method of obtaining the correct solution (Wagner & Sternberg, 1985). Note that these characteristics do not apply as well to many of the problems people face in their daily lives, including many of the problems at work. In direct contrast, work problems [yardstick for “practical intelligence”] often are (a) unformulated or in need of reformulation, (b) of personal interest, (c) lacking in information necessary for solution, (d) related to everyday experience, (e) poorly defined, (f) characterized by multiple “correct” solutions, each with liabilities as well as assets, and (g) characterized by multiple methods for picking a problem solution (p. 913).
	<i>Test design fallacy #2</i>	<i>Intelligence is marble collection</i>	<i>Portraying general intelligence (g) as if it were just an aggregation of many separate specific abilities or skills, not a singular phenomenon in itself (Entry 10), because IQ batteries calculate IQs by adding up scores on different subtests (Entry 9).</i>

iv	Flynn (2007). From example ii above	Context: Author is proposing a skills-based definition of intelligence that is “narrow enough to offer good advice to those who want to make intelligence measurable and specific” (p. 55).	Quotation from Example ii above: The subtest...Arithmetic measures learning what schools teach as mathematics. Comprehension measures understanding the mechanics of everyday life. Coding and Symbol Search measure processing speed. Forward Digit Span isolates memory from the other components [other individual marbles] of intelligence broad [the collection of marbles] (p. 55).
v	Flynn (2007)	Context: Author is explaining how secular increases in IQ test scores can represent a rise in overall intelligence but not in <i>g</i> , the issue at hand being that scores on some highly <i>g</i> -loaded IQ subtests (e.g., Similarities) have risen a lot but others (e.g., Vocabulary) hardly all—“or, how can IQ gains be so contemptuous of <i>g</i> loadings?” (p. 9).	Quotation: My fundamental line of argument will be that understanding intelligence is like understanding the atom: we have to know not only what holds its components together but also what splits them apart. What binds the components [marbles] of intelligence [the collection] together is the general intelligence factor or <i>g</i> ; what acts as the atom smasher is the Flynn effect or massive IQ gains over time (p. 4)...At any particular time, factor analysis will extract <i>g</i> (IQ)—and intelligence [the collection] appears unitary. Over time, real-world cognitive skills [individual marbles] assert their functional autonomy and swim freely of <i>g</i> —and intelligence appears multiple (p. 18)...Asking whether IQ gains are intelligence gains is the wrong question because it implies all or nothing cognitive progress. The twentieth century saw some cognitive skills [marbles] make great gains, while others were in the doldrums. To assess cognitive trends, we must dissect “intelligence” [the collection] into solving mathematical problems, interpreting the great works of literature, finding on-the-spot solutions, assimilating the scientific worldview, critical acumen, and wisdom [individual marbles] (p. 10).
vi	Howe (1997)	Context: Author is listing “Twelve Well-Known ‘Facts’ about	Quotation: 8. An IQ test score is no more than an indication of someone’s performance at a range of mental tasks. The implication that there is just one all-important dimension of intelligence is wrong and unhelpful. Other

		Intelligence Which are Not True” (p. 161)	kinds of intelligence [marbles] can be equally crucial (p. 162).
	<i>Score variation fallacy #1</i>	<i>Non-fixedness proves malleability</i>	<i>Using evidence of any fluctuation or growth in the mental functioning of individuals as if it were proof that their rates of growth can be changed.</i>
vii	World News Tonight with Peter Jennings (1994)	Context: Newscaster is contesting <i>The Bell Curve</i> 's claim that intelligence is a stable, measurable trait.	Quotation: BETH NISSEN: <i>[voice-over]</i> ...Using high-tech scanners and imagers, neuroscientists like Dr. Eric Kandel can actually see why intelligence is almost impossible to measure—it is constantly changing [non-fixedness]. The brain, the factory that produces intelligence, is always learning, retooling. Dr. ERIC KANDEL: You can actually show an anatomical change; an actual increase in the number of synaptic connections [non-fixedness]. BETH NISSEN: <i>[voice-over]</i> Brain surgeons like Dr. Benjamin Carson say the brain responds to everything it experiences, from its first formation in utero [non-fixedness]. Dr. BENJAMIN CARSON, Johns Hopkins University: I would have to say that hydration, nutrition and stimulation, environmentally, play very large roles in the development of the human brain [non-fixedness]. BETH NISSEN: <i>[voice-over]</i> That challenges the most critical and criticized claim in the new book that while environment may have an effect, intelligence is largely genetic and largely fixed in a person by the age of 16 or 17 (p. 1) [rebutting the straw man that genetic means everything about the brain and intelligence is “fixed” in stone by age 16 or 17]
viii	Howe (1997)	Context: Author is discussing what he considers better alternatives to “traditional intelligence theory.”	Quotation: [These newer approaches] acknowledge that human intelligence is far from fixed, and that it is subject to development processes [non-fixed]. [For example], Anderson is aware that despite the fact that the contents of intelligence tests administered to young children are very different from those of adult tests, intelligence theory has largely ignored the fact that human intelligence develops [supposed blindness to non-

			<p>fixedness] rather than being static [supposed belief in fixedness]. Anderson’s approach is intended to remedy this situation. However, since he wishes to retain some aspects of the g concept, which is essentially unchangeable [non-malleable] by definition, in order to make allowance for the fact that intelligence does nevertheless develop he is forced to include in his model both developing and unchanging elements (p. 138).</p>
	<i>Score variation fallacy #2</i>	<i>Improvability proves equalizability</i>	<i>Using evidence that intellectual skills and achievements can be improved within a population as if it were proof that they can be equalized in that population.</i>
ix	Howe (1997)	Context: Author is arguing for interventions to raise IQs in disadvantaged groups.	Quotation: There exists a large amount of convincing evidence that a person’s intelligence level can alter, sometimes very substantially [improvability]In a prosperous society, only a self-fulfilling prophecy resulting from widespread acceptance of the false visions expounded by those who refuse to see that intelligence is changeable would enable the perpetuation of a permanent caste of people who are prevented from acquiring the capabilities evident in successful men and women and sharing their rewards [equalizability] . Unfortunately, however, at present just that set of circumstances appears to be in place. Underclasses do not emerge for no reason; they are created by unequal societies (pp. 62-63).
x	The White House (2001)	Context: Executive Summary of No Child Left Behind Act of 2001 on White House website is highlighting intent to close achievement gaps by bringing all students up to the same high level of achievement.	<p>Quotation:</p> <p>Closing the Achievement Gap: [equalizability]</p> <ul style="list-style-type: none"> • Accountability and High Standards. States, school districts, and schools must be accountable for ensuring that all students, including disadvantaged students, meet high academic standards. States must develop a system of sanctions and rewards to hold districts and schools accountable for improving academic achievement [improvability]. <p>....</p> <p>Rewarding Success and Sanctioning Failure:</p> <ul style="list-style-type: none"> • Rewards for Closing the Achievement Gap. High performing states that narrow the achievement gap [equalizability] and improve

			overall student achievement will be rewarded [improvability] .
xi	Dionne (1994)	Context: <i>Washington Post</i> columnist is arguing that <i>The Bell Curve</i> “is not a ‘scientific’ book at all but a political argument offered by skilled polemicists aimed at defeating egalitarians.”	Quotation: If you had any doubts that we live in a time of deep pessimism about the possibility of social reform, the revival of interest in genetic explanations for human inequality ought to resolve them.... Whenever the social reformers are seen as failing, along come allegedly new theories about how the question for greater fairness or justice or equality [equalizability] is really hopeless because people and groups are, from birth, so different, one from another....That is the real significance of the appearance of and interest in” <i>The Bell Curve</i> ”...The implicit argument of the book is that if genes are so important to intelligence and intelligence is so important to success, then many of the efforts made over the past several decades to improve people’s life chances [improvability] were mostly a waste of time. Herrnstein and Murray never quite say that.
	<i>Score variation fallacy #3</i>	<i>Interactionism (gene-environment co-dependence) nullifies heritability</i>	<i>Portraying the gene-environment partnership in creating a phenotype as if conjoint action within the individual precluded teasing apart the roots of phenotypic differences among individuals.</i>
xii	Sternberg (1997)	Context: Author is distinguishing “conventional IQ-based view” of intelligence from his proposed “successful intelligence.”	Quotation: Intelligence is partially heritable and partially environmental, but it is extremely difficult to separate the two sources of variation, because they interact in many different ways [interactionism] . Trying to assign an average number to the heritability of intelligence is like talking about the average temperature in Minnesota (p. 48).
xiii	Andrews & Nelkin (1996)	Context: Letter to <i>Science</i> is disputing conclusions in <i>The Bell Curve</i> .	Quotation: As geneticists and ethicists associated with the Human Genome project, we deplore <i>The Bell Curve</i> ’s misrepresentation of the state of genetic knowledge in this area....First, Herrnstein and Murray invoke the authority of genetics to argue that “it is beyond significant technical dispute that cognitive ability is substantially heritable.”...Many geneticists have pointed out the enormous scientific and methodological problems in attempting to separate genetic components from environmental

			contributors, particularly given the intricate interplay between genes and the environment [interactionism] that may affect such a complex human trait as intelligence (p. 13).
	<i>Score variation fallacy #4</i>	<i>99.9% similarity negates differences</i>	<i>Portraying the study of human genetic variation as irrelevant or wrong-headed because humans are 99.9% (or 99.5%) alike genetically, on average.</i>
xiv	Park (2002)	Context: Anthropology textbook is discussing “why there are no biological races within the human species” (p. 396).	Quotation: The nonexistence of definable [biological] racial groups coincides with and reinforces our ethical ideas of human equality [no races would be a more ethical empirical fact]. But wishful thinking cannot take the place of scientific rigor. We must be able to say <i>why</i> there are no races....We need to present sound scientific evidence for it (p. 395)....What do [the genetic data] tell us? When comparing any two humans, it looks as if only, at most, about 3 million of our 3 billion nucleotides are SNPs [differences in the genome at the level of base pairs]. In other words, any two humans differ genetically by less than one-tenth of one percent (0.1 percent) [99.9% alike genetically]....All the phenotypic variation that we try to assort into race is the result of a virtual handful of alleles [fraction of 3 million SNPs = trivial difference] (pp. 397-398).
xv	Holt (1994)	Context: <i>New York Times</i> Op-Ed is disputing the idea that racial differences in intelligence could have any genetic basis.	Quotation: [G]enetic diversity among the races is minuscule [near irrelevance]. Molecular biologists can now examine genes in different geographical populations. What they have found is that the overwhelming majority of the variation observed—more than 85 percent—is among individuals within the same race. Only a tiny residue [near irrelevance] distinguishes Europeans from Africans from Asians.
xvi	Marks (1995)	Context: Author is summing up his book’s argument that genetic differences by race are minor but exaggerated in order	Quotation: The categories we acknowledge as races are marked by any number of differences, but the biological differences between them are minimal [near irrelevance], reinforced by social and cultural difference (pp. 274-275)....Providing explanations for social inequalities as being rooted in nature is a classic pseudoscientific occupation [wrong-headed]. It has always been welcome, for it provides those in power with a natural

		to justify and perpetuate social inequality.	validation of their social status (p.273).
	<i>Test validation fallacy #1</i>	<i>Contending definitions negate evidence</i>	<i>Portraying lack of consensus in verbal definitions of intelligence as if that negated evidence for the construct validity of IQ tests.</i>
xvii	Singham (1995)	Context: Author is advising educators that <i>The Bell Curve</i> is unscientific and ideological.	Quotation: Intelligence is an elusive concept. While each person has his or her own intuitive methods for gauging the intelligence of others [lack of consensus], there is no a priori definition of intelligence that we can use to design a device to measure it (p. 272)... [implication: results from existing devices may be ignored] All kinds of hypotheses can be invoked to explain the data [showing correlations among intelligence, race, and socioeconomic status]. And this shouldn't be too surprising. As I emphasized above, both race and intelligence are poorly defined and operationally ambiguous. When you have two variables that are ill-defined, it is asking too much to expect a simple relationship between them to emerge (p. 278).
xviii	“The Bell Curve Agenda” (<i>New York Times</i> , 1994)	Context: Editorial is arguing that “what is new about [<i>The Bell Curve</i> book]—the fixation on genes as destiny—is surely unproved and almost surely wrong” and therefore IQ level actually is manipulable.	Quotation: There is wide agreement among researchers on intelligence that genetic inheritance influences mental acuity, but there is also wide agreement that life experiences, even in the womb, exert a powerful influence on brain structure. Further, there is wide disagreement about what intelligence consists of and how — or even if — it can be measured in the abstract [lack of consensus]. For example, in “The Mismeasure of Man,” Stephen Jay Gould, the evolutionary biologist, dismissed “the I.Q. industry” as little more than an effort by men of European descent to maintain their prominence in the world (p. A16) [implication: test results represent social privilege].
	<i>Causal fallacy #1</i>	<i>Phenotype is genotype</i>	<i>Portraying phenotypic differences in intelligence (Entry 5) as if they were necessarily genotypic (Entry 1).</i>
xix	Duster (1995)	Context: Author	Quotation: Those making the claims about the genetic component of an

		arguing that “there has always been a tendency to link existing social orders with so-called innate physical, intellectual and spiritual qualities.”	array of behavior and conditions (crime, mental illness, alcoholism, gender relations, intelligence) come from a wide range of disciplines....Richard Herrnstein (1971), the late Harvard psychologist not only argued the genetics of intelligence but even speculated that someday “the tendency to be unemployed may run in genes.” And it is sociologist, Robert Gordon (1987), who argues that race differences in delinquency are best explained by IQ differences between the races, not socioeconomic status (p. 1) [Gordon’s claim about phenotypic group IQ differences is treated as if a genetic claim] [Note: This example also conflates claims about differences <i>within</i> a race (Herrnstein’s concern) with claims about average differences <i>between</i> races (Gordon’s concern) in order to impugn the latter.]
xx	Bartholomew (2004)	Context: Author is describing the difficulty of determining whether the black-white IQ difference originates in whole or part in the genes or whether it can be wholly accounted for by environmental factors (p. 122).	Quotation: In order to resolve the uncertainty about how to interpret this [black-white IQ] difference it was, and <i>is</i> , necessary to do two things. First, to demonstrate whether the difference is really due to some environmental factor that is confounded with race. Secondly to identify a relevant genetic difference between the two groups, assuming one exists. The possibility of confounding has given rise to an enormous amount of work. Often this is spoken of under the heading of test bias [is the measured IQ difference really an intelligence difference?] . A test is biased if it gives an advantage to one group rather than the other. In other words, we cannot be sure whether the score difference is due to ability to do the test or to environmental factors which affect the groups differently [unclear which question being addressed—are IQ scores biased measures of black intelligence? vs. are validly measured black-white differences in intelligence environmentally caused?] . This is often described in terms of cultural differences. As with the smoking and cancer example used above, one can never absolutely rule out environmental explanations of this kind [what causes real differences in health?] (pp. 122-123).
	<i>Causal fallacy #2</i>	<i>Biological is genetic</i>	<i>Portraying biological differences (such as brain phenotypes, Entry 4) as if they were necessarily genetic (Entry 1).</i>

xxi	Bartholomew (2004)	Context: Author is discussing possible sources of Flynn Effect (average IQ is rising).	Quotation: At first sight one might see this [extraordinary secular increase in IQ] as very strong empirical evidence for the determination of IQ by environmental factors because it is difficult to see what biological factors [biological vs. environmental, as if biological=genetic] could do so much in so little time. Equally however, and given our knowledge of the modest effects that environmental factors typically have, it is not easy to imagine what environmental factors could produce such a big change in such a relatively short time. [Thus w]hatever has happened cannot reasonably be attributed to the additive effects of heredity and environment (p. 138) [genetic vs. environmental factors] .
xxii	News and Notes (NPR, 2007)	Context: NPR is following up an interview with J. P. Rushton, who spoke about the correlations between race, brain size, and intelligence, by interviewing a critic of intelligence research.	Quotation: [Farai Chideya]: Why don't you talk to us a little bit about this issue of brain size and intelligence? Do you see any link? [Rushton] says that it is absolutely incontrovertible that there is a link. What's your research or what does research that you've looked at tell you? [Bill Tucker]: Well, there are many criticisms of the studies on brain size and intelligence, but quite apart from the scientific issues I think that there are some obvious practical facts that would suggest that this link is not as firm as Rushton claims it is. For example, one of the individuals who is usually proclaimed as one of the most intelligent persons of the 20 th century, Albert Einstein, left his brain to science. It was studied. It is slightly below average for his size....So to suggest that brain size is linked to intelligence when one of the most intelligent persons ever had a below average brain size would suggest that there are serious doubts about this work. [invokes imperfect correlatio to ignore the .4 correlation between in vivo brain size and intelligence, presumably because biological differences implicate genetic ones]
	<i>Causal fallacy #3</i>	<i>Environment is nongenetic</i>	<i>Portraying external environments (Entry 3) as if they were necessarily nongenetic, that is, unaffected by and unrelated to the genotypes of individuals in them.</i>
xxiii	Monastersky (2008)	Context: News article	Quotation: For generations, psychologists have noted that children raised in

		<p>is reporting research on about “how poverty alters the brain.”</p>	<p>poverty perform poorer on cognitive tests, on average, than do students from wealthier families. Some researchers have taken those results to argue that intelligence is determined for the most part by genetics and that certain races are inherently smarter than others [*]....But the new results from neuroscience indicate that experience, especially being raised in poverty, has a strong effect on the way the brain works. “It’s not a case of bad genes,” said Ms. Farah [but the study did not consider or control for genetic differences among either parents or children]...The researchers studied a group of African-American children of low socioeconomic status, who had been tracked from birth through high-school.... [MRI scans showed that] the students raised in more nurturing homes had bigger hippocampi, the portion of the brain associated with forming and retrieving memories....In [another] study, researchers put a net of electrodes on the heads of children and measured their brain waves. The children were seated between two speakers playing different stories and they were asked to pay attention to only one of the stories. While the stories were being read, the children heard identical bursts of distracting noise coming from either of the speakers....The study revealed that students from lower-income families were less able to screen out the noises embedded in the stories they were supposed to ignore....With those results and others suggesting that cognitive skills are strongly influenced by environment [but only if one ignores the usual genetic correlations between parental intelligence and income and between parent and child intelligence], the Oregon team is developing intervention programs to try to counteract the effects of poverty. [*Note: Here, the article is committing the Phenotype-Is-Genotype Fallacy but attributing it to the unnamed “some researchers.”]</p>
xxiv	Fischer et al. (1996)	Context: Authors are arguing that the AFQT measures differences in	Quotation: What [the AFQT] captures best is how much instruction people encountered and absorbed. It does that better than does the conventional “years of education” measure, because the AFQT seems to assess educational quality and information instruction as well as simply time in

		opportunity to learn, not in “raw intelligence.”	school. It taps the differences between those who spent time in classes with rich curricula, energetic teachers, motivated students, and plentiful resources and those who spent time in classes without those qualities. It taps the difference between those who are “instructed” outside the classroom and those who are not.Another way to understand what we have shown is that test takers’ AFQT scores [cognitive performances] are good summaries of a host of prior experiences (mostly instruction) [external environments] that enable someone to do well in adult life (p. 68).
	<i>Standard of evidence fallacy #1</i>	<i>Imperfect measurement pretext</i>	<i>Maintaining that valid, unbiased intelligence tests should not be used for making decisions about individuals until the tests are made error-free.</i>
xxv	FairTest (2007)	Context: University Testing Fact Sheet on FairTest website is arguing that the ACT, SAT, and SAT Subject Tests are not accurate enough to be used in evaluating applicants for college admissions and scholarships.	Quotation: ACT scores are imprecise. The individual tests have large margins of error, according to data from ACT. The margin of error - the inconsistency in ACT scores inherent in the testing process - on each subject's 1-36 point scale is 1.55 points in English, 1.43 in Mathematics, 2.20 in Reading, and 1.75 in Science Reasoning. In other words, if a student were to retake the exam, there would be about a two-thirds chance that her score would be 1.55 points higher or lower on the English test than on a previous administration of the test. There is also a one-third chance the score difference would be even larger [appeals to imperfection]. The margins of error, while appearing to be small at 1.43 - 2.20, can actually have significant consequences for applicants when admissions offices or financial aid programs require minimum (or "cut-off") scores.... The ACT's flaws have serious consequences [imperfection is harmful]. Despite its inaccuracies, biases, and coachability, ACT cut-off scores are often used to determine entrance into schools and allocate scholarships....The weak predictive power of the ACT, its susceptibility to coaching, examples of test score misuse, and the negative impact test score use has on educational equity all lead to the same conclusion - test scores should be optional in college admissions [call to reduce testing].
xxvi	Miller (2001)	Context: News article	Quotation: Scholars agree with educators and policymakers that tests are

		in <i>Chronicle of Higher Education</i> reporting complaints in education profession about large-scale testing.	useful for tracking children's progress and identifying weaknesses in teaching. But Mr. Valencia and other education researchers have begun describing testing's dark side [imperfection is harmful]. Standardized tests, they say, are too limited, too imprecise, and too easily misunderstood to form the basis of crucial decisions about students [call to reduce testing].... For one thing, tests are imprecise yardsticks of a student's abilities [appeal to imperfection]. Ideally, a child would earn the same score on variations of the same test given on different days. (Psychometricians would say such a test had a reliability of 100 percent.) But that threshold is beyond reach. Students' scores vary from day to day, depending on their health, their mood, or even what they ate for breakfast. Furthermore, it's difficult to keep exams consistent from year to year. Test designers must constantly refresh the test questions, but the new items are never precisely comparable to the old ones. That's why designers publish the margins of error of their products, expressed as "reliability coefficients" between 0 and 1. Most standardized tests used to evaluate elementary and secondary students claim a reliability coefficient in the neighborhood of .9, "plenty good for most purposes," says David R. Rogosa, a professor of education at Stanford University and an expert in educational assessment. "But a reliability of .9 ain't all it's cracked up to be" (p. A14).
xxvii	Hartigan & Wigdor (1989)	Context: National Academy of Sciences (NAS) report is explaining why it is recommending that the US Employment Service (USES) continue to race-norm job applicants' employment test scores.	Quotation: In sum, the modest validities of the GATB cause selection errors [appeal to imperfection] that weigh more heavily on minority workers than on majority workers [because the rate of false rejections is higher in any lower-scoring group, regardless of race]. This outcome is at odds with the nation's express commitment to equal employment opportunity for minority workers [suggests social harm]. In the committee's judgment, the disproportionate impact of selection error provides scientific grounds for the adjustment of minority scores so that able minority workers have approximately the same chances of referral as able majority workers....The committee has analyzed two score-adjustment methods—the current USES system of within-group percentile scores and a

			performance-based method of computing scores. Both score adjustment strategies are race-conscious [introduce error in form of racial bias]; both would virtually eliminate the adverse impact of the GATB [General Aptitude Test Battery] on black and Hispanic subpopulations...and both adjustments would be commensurate with the far less than perfect relation between the GATB test score and job performance [appeal to imperfection] (pp. 7-8). [Note: USES eliminated the GATB when it could not longer race-norm it.]
	<i>Standard of evidence fallacy #2</i>	<i>Dangerous thoughts trigger</i>	<i>Maintaining that scientific conclusions purported to be dangerous or divisive should not be entertained until proved beyond all possible doubt.</i>
xxviii	C. Kiesler (January 17, 1980, personal communication to A. R. Jensen)	Context: Editor of the <i>American Psychologist</i> is explaining why he is rejecting Arthur Jensen's manuscript, "The Nature of the Average Difference between Whites and Blacks on Psychometric Tests: Spearman's Hypothesis" (which was later published as a target article in <i>Behavioral and Brain Sciences</i> , 1985, 8, 193-219).	Quotation: My own feeling as Editor is that since this area is so controversial and important to our society, I should not accept any manuscript that is less than absolutely impeccable. I do have some serious doubts and reservations about this analysis and these data. In this paper there is a hanging implication that any differences that are demonstrated to exist are genetic [the dangerous idea]. Therefore one has to look at the statistical procedures and the definitional process very thoroughly to assure one's self that other [nongenetic] possibilities are not possible or plausible (p. 1). [Note: Spearman's hypothesis is about phenotypic differences, not genetic ones.]
xxix	Hunt & Carlson (2007)	Context: Authors are proposing standards for conducting and	Quotation: Scientists cannot be held responsible for the use that others make of information they provide. They can be held responsible for stating the quality of the information they provide and for presenting alternative

		evaluating research on group differences in intelligence.	interpretations of that information when appropriate. On a topic as divisive as racial/ethnic differences in intelligence, this is a very serious issue. We do not see any need for [Jensen’s] potentially divisive “default hypothesis” [that the causes of individual and group differences are the same] emphasizing either biological or social factors [the dangerous idea], in the absence of convincing evidence that rules out other hypotheses [proof beyond all possible doubt] (p. 210).
	<i>Standard of evidence fallacy #3</i>	<i>Happy thoughts leniency</i>	<i>Maintaining that mere theoretical possibility elevates the scientific credibility of a politically popular idea above that of an empirically plausible but unpopular conclusion.</i>
xxx	Diamond (1999)	Context: Author is arguing that “biological differences” cannot account for “why...human development proceed[ed] at such different rates on different continents” over human history, despite seemingly compelling arguments that they do (p. 16).	Quotation: A seemingly compelling [empirically plausible] argument goes as follows. White immigrants to Australia built a literate, industrialized, politically centralized, democratic state based on metal tools and on food production, all within a century of colonizing a continent where the Aborigines had been living as tribal hunter-gathers without metal for at least 40,000 years. Here were two successive experiments in human development, in which the environment was identical and the sole variable was the people occupying that environment. What further proof could be wanted to establish that the differences between Aboriginal Australian and European societies arose from differences between the peoples themselves? The objection to such racist explanations is not just that they are loathsome, but also that they are wrong. Sound evidence for the existence of human differences in intelligence that parallel human differences in technology is lacking. In fact, as I shall explain in a moment, modern “Stone Age” peoples are on the average probably more intelligent, not less intelligent, than industrialized peoples [the theoretically possible] (p. 19).
xxxi	“The Bell Curve Agenda” (<i>New York Times</i> , 1994)	Context: Editorial is arguing that “what is new about [<i>The Bell Curve</i> book]—the fixation on genes as	Quotation: “The Bell Curve” presumes, but does not prove, that differences in genes account for 60 percent of the differences in the I.Q.’s of children. It is essential to note—which the authors do but many of their critics do not—that <i>group</i> differences in I.Q. may have nothing to do with genes even if <i>individual</i> I.Q.’s are largely inherited. An example proves the point.

		<p>destiny—is surely unproved and almost surely wrong” and therefore IQ level actually is manipulable.</p>	<p>Plants grown together under ideal conditions [theoretically possible but implausible for humans] will achieve different heights based solely on individual genetic makeup. But lock half the plants in a dark closet [also theoretically possible but totally implausible for humans] and the difference in average height of the two groups will be due entirely to environment [under these totally implausible conditions]. So even if I.Q.’s are deemed to be largely inherited that says nothing about the potential [theoretically possible] impact on I.Q. of altering prenatal care or aggressive early education (p. A16).</p>
--	--	--	--