Four Empirically-Based Reasons Not to Administer Time-Limited Tests

Morton Ann Gernsbacher, Ph.D.
University of Wisconsin-Madison
magernsb@wisc.edu

Raechel N. Soicher, M.A.
Oregon State University
raechel.soicher@oregonstate.edu

Kathryn A. Becker-Blease, Ph.D.
Oregon State University
kathryn.blease@oregonstate.edu

Corresponding Author: Morton Ann Gernsbacher, PhD
Vilas Research Professor & Sir Frederic Bartlett Professor
Department of Psychology
1202 W. Johnson St.
University of Wisconsin-Madison
Madison WI 53706
(608) 262-6989
magernsb@wisc.edu

© 2020, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/tps0000232
Abstract

For more than a century, measurement experts have distinguished between time-limited tests and untimed power tests, which are administered without time limits or with time limits so generous all students are assured of completing all items. On untimed power tests, students can differ in their propensity to correctly respond to every item, and items should differ in how many correct responses they elicit. However, differences among students’ speed of responding do not confound untimed power tests; therefore, untimed power tests ensure more accurate assessment. In this article, we present four empirically-based reasons to administer untimed power tests rather than time-limited tests in educational settings. (1) Time-limited tests are less valid; students’ test-taking pace is not a valid reflection of their knowledge and mastery. (2) Time-limited tests are less reliable; estimates of time-limited tests’ reliability are artificially inflated due to artefactual consistency in students’ rate of work rather than authentic consistency in students’ level of knowledge. (3) Time-limited tests are less inclusive; time-limited tests exclude students with documented disabilities who, because they are legally allowed additional test-taking time, are often literally excluded from test-taking classrooms. (4) Time-limited tests are less equitable; in addition to excluding students with documented disabilities, time-limited tests can also impede students who are learning English, students from under-represented backgrounds, students who are older than average, and students with disabilities who encounter barriers (e.g., stigma and financial expense) in obtaining disability documentation and legally-mandated accommodations. We conclude by offering recommendations for avoiding time-limited testing in higher educational assessment.

Keywords: timed testing, disability, inclusion, equity, higher education

Public Significance Statement: This article cautions instructors against administering tests that have time limits because time-limited tests are less valid, less reliable, less inclusive, and less equitable. Instead, instructors are encouraged to administer untimed power tests, which are tests without any time limits or with time limits so generous all students are assured of completing all items.
“Yeah, everywhere around the country. What happened is, all the wealthy families ... figured out that if I get my kid tested and they get extended time, they can do better on the test”

[Cooperating Witness-1, from United States Department of Justice, 2019, p. 25]

For more than a century, measurement experts have distinguished between time-limited tests and untimed power tests (Lorge, 1936; McCall, 1916; Ruch & Koerth, 1923). Power tests are typically administered without any time limits or with time limits so generous all students are assured of having plenty of time to complete every item (Sisk, 1926). Students taking an untimed power test will most likely differ in their propensity (or preparation) to correctly respond to every item, and items on an untimed power test should be designed to differ in how many correct responses they elicit (Kyllonen & Zu, 2016). However, differences among students’ speed of responding do not confound untimed power tests; therefore, untimed power tests ensure more accurate assessment (Colker, 2019; Donlon, 1975).

Untimed power tests have been regarded, for nearly a century, as “the ideal” (Peak & Boring, 1926, p. 72) because they best assess the “complexity of … thought-processes” (Line & Kaplan, 1932, p. 1). In contrast, time-limited tests are good only for assessing tasks that are so cognitively simple that everyone would respond correctly if no time limit was imposed (Danthiir, Wilhelm, & Schacht, 2005; DuBois, 1932; Mead & Drasgow, 1993). For example, on the Differential Aptitude Tests only lower-level clerical speed and accuracy are measured with time limits; higher-level abstract reasoning, verbal analysis, spatial relations, and mechanical thought are assessed via untimed power tests (Bennett, Seashore, & Wesman, 1956).

In this article, we present four empirically-based reasons not to administer time-limited tests in higher educational settings and to instead administer untimed power tests. Our reasons are that time-limited tests are less valid, reliable, inclusive, and equitable.

**Time-Limited Tests Are Less Valid**

For nearly a century, we have known that students’ pace on an untimed power test does not validly
reflect their performance. Some students work quickly but perform poorly; other students work slowly but perform well, and all variations exist in between (Mudge, 1921). The lack of prediction from pace to performance has been documented across a variety of untimed power tests and a wide range of large student samples, including grade-school and high-school students (Dodonova & Dodonov, 2013; Freeman, 1923; Line & Kaplan, 1932; Mangan, 1959; Slater, 1938; Terranova, 1972; Vernon & Kantor, 1986), as well as college students (Dodonova & Dodonov, 2013; DuBois; 1932; Freeman, 1923; 1932; Lord, 1956; Miller & Weiss, 1976; Neubauer, 1990; Tinker, 1944; Tryon & Jones, 1933; White, 1973; Yates, 1963, 1966).

In the classroom, one can observe the poor prediction of test-taking pace to test-taking performance by simply correlating the order in which students finish an exam with the grade they earn on that exam. Indeed, nearly a century ago, two psychology professors did just that. Longstaff and Porter (1928, p. 638) described their motivation in the following way: “We hear school men very authoritatively saying that the fast students make the best grades and the slow ones the poorest. … Statements of this kind are usually based on the assumption that if a student knows the subject in which he is being tested it should follow that he requires but a short time to make his answer. Needless to say, this assumption merits confirmation.” Yet, confirmation was not obtained. Rather, Longstaff and Porter’s (1928) data, based on nearly 200 General Psychology students each taking nearly 15 exams, demonstrated “very little, if any, relationship between the times required and the scores made” (p. 641).

Forty years later, Burack (1967, p. 164) reminisced that “psychology majors in the 1930s” had read studies demonstrating the poor prediction of test-taking pace to test-taking performance in psychology courses and wondered if the prediction still failed; it did (see also Michael & Michael, 1969). Across the next thirty years, the prediction continued to fail in psychology courses across the country (Becker & Suls, 1982; Bridges, 1985; Ebel & Frisbie, 1991; Foos, 1989; Herman, 1997; Johnston, 1987; Kennedy, 1994; Lester, 1991; Paul & Rosenkoetter, 1980; Wierzbicki, 1994).

Into the twenty-first century, the prediction continues to fail not only in psychology courses (Brothen, 2012; Hammonds & Mariano, 2015) but also in chemistry (Ghaffari, 2015; Nevo & Spector,
1979); engineering (Armitage, 1999); agriculture (McDannell & Lyvers Peffer, 2013); business, including finance, real estate, and management (Beaulieu, & Frost, 1994; Lee, 2012; Schnusenberg & Slater, 2011); English and literature (Lovett, Lewandowski, & Potts, 2017; Rogers, 1968); statistics (Hammonds & Mariano, 2015; Onwuegbuzie & Daley, 1996); and mathematics (Rogers, 1968). Why does the prediction fail? The prediction fails because putting time limits on power tests introduces irrelevant variance (Ackerman & Ellingsen, 2016; Donlon, 1975; Lu & Sireci, 2007; Peterson, 1993; Pitoniak & Royer, 2001; Thissen, 1983).

The empirical evidence against the assumption that “fast students make the best grades” comes from thousands of students taking all types of tests: graduate as well as undergraduate examinations (Freeman, 1923; Michael & Michael, 1969; Onwuegbuzie & Daley, 1996); open book as well as closed book exams (Michael & Michael, 1969); and exams comprising problem sets (Armitage, 1999), essays (Burack, 1967; Foos, 1989; Lee, 2012), short-answer (Burack, 1967; Onwuegbuzie & Daley, 1996; Wierzbicki, 1994), as well as multiple-choice items (Burack, 1967; Bridges, 1985; Foos, 1989; Freeman, 1923; Longstaff & Porter, 1928; Paul & Rosenkoetter, 1980; Schnusenberg & Slater, 2011; Wierzbicki, 1994).

Why does this erroneous assumption persist? It persists from faulty reasoning and misaligned analogies. For example, a recent OpEd likened professors providing extended time on classroom exams, which is a topic we address later, to World Track and Field officials providing Usain Bolt’s competitor a “a 20-metre head start” (Pardy, 2017). But exams are not track meets, and we have known for more than a century, as Whipple documented in his classic Manual of Mental and Physical Tests (1914, p. 10), that “if we seek to evaluate the complex ‘higher’ mental functions, speed is not the primary index of efficiency, as is borne out by the evidence that speed and intelligence are not very highly correlated.”

We have also known for nearly a century that beyond classrooms and into occupations, skill also uncouples from pace. For example, Lemmon (1927) reported a poor correlation between trained pilots’ response speed and their aviation ability, akin to the poor correlation between students’ test-taking pace and their test-taking performance. Berliner (1994) identified professional tasks on which novice teachers
required more time but other tasks on which experienced teachers required more time. Similarly, because more knowledgeable students do not take less time on time-limited tests and tests become less valid as they tap more intellectual content (Whipple, 1914), we recommend untimed power tests in higher education settings.

**Time-Limited Tests Are Less Reliable**

For almost as long as measurement experts have distinguished time-limited tests from untimed power tests, experts have cautioned about the spurious reliability of time-limited tests (Angoff, 1953; Cooper, 1984; Cronbach & Warrington, 1951; Thissen, 1983; Wesman, 1949). For example, in their classic textbook, *Introduction to Classical and Modern Test Theory*, Crocker and Algina (1986) warned that reliability estimates of time-limited tests are artificially inflated due to artefactual consistency in students’ rate of work rather than authentic consistency in students’ level of knowledge. Thus, just as variation between students’ time-limited performance confounds estimates of time-limited tests’ validity, consistency within students’ time-limited performance inflates estimates of time-limited tests’ reliability (Cronbach, 1949).

Anne Anastasi, psychology’s “test guru” (Goode, 2001) issued more dire warnings, noting “It has been repeatedly demonstrated, both theoretically and empirically,” that the reliability coefficients of time-limited tests “may be completely meaningless” when computed using standard procedures (Anastasi & Drake, 1954, p. 529). Although some statistical corrections for computing the reliability of time-limited tests have been suggested, we have known for decades that the best way to improve a time-limited test’s reliability is simply to remove its time limits (Carter, 1932; Traub & Hambleton, 1972).

We suppose most instructors would like to administer untimed tests. We assume most instructors want to assess students’ level of acquired knowledge (DuBois, 1932; Tinker, 1944) not their speed in reacting to test items (Mead & Drasgow, 1993; Peak & Boring, 1926). We believe most instructors prize genuinely reliable untimed power tests over artificially reliable time-limited tests, although some instructors continue to erroneously conflate speed with power. Yet, we observe many instructors inadvertently

As Morrison (1960, p. 231) wrote decades ago, “the popularity of time-limit tests is due more to their practical administrative advantages … than to any experimentally-supported rationale governing the imposition of time limits on performance.” Class periods are finite; even the typically more generous final exam periods are bound. These time limits introduce a “spurious … contribution” (Wilhelm & Schulze, 2002, p. 537) that can “skew measures” (Henderson, 2005) to produce “severe problems” (Peterson, 1993, p. 9) and alter the “construct the test intends to measure” (Lu & Sireci, 2007, p. 31). Therefore, we reiterate Mollenkopf’s (1960, p. 223) decades-old admonition: “you should have a better reason for setting whatever [time] limits you impose than mere convenience of administration.”

**Time-Limited Tests Are Less Inclusive**

The faulty assumption that “fast students make the best grades and the slow ones the poorest” has also disparaged expectations about students with disabilities. For decades, *slow* was a euphemism for *intellectual disability* (Abrams & Goodman, 1998). The late twentieth-century term *mental retardation*, which referred to what we now call *intellectual disability*, meant “an abnormal slowness of thought or action” (Merriam-Webster, no date) and was derived from the Latin term *retardâre* (“to make slow,” The American Heritage Dictionary, no date). An early, but mistaken, view of intelligence promoted this erroneous assumption: “To be quick is to be intelligent, while slowness is synonymous with dullness” (Line & Kaplan, 1932, p. 1).

Even in the twenty-first century this unfortunate assumption persists, as demonstrated by the 2019 college entrance scandal, during which affluent parents illegally obtained college admission for their children and from which we quoted at the outset of our article. One parent was instructed by the alleged fixer
“to tell [their daughter] when she gets tested, … to be stupid, not to be as smart as she is. The goal is to be slow, to be not as bright” (U.S. Department of Justice, 2019, p. 25, emphasis added).

One of the loopholes exploited in the college entrance scandal capitalized on U.S. laws that protect the rights of students with disabilities, including the Rehabilitation Act of 1973, in particular, Section 504, and the Americans with Disabilities Act of 1990, in particular, Section 309, which extended disability anti-discrimination protections to the private sector (United States Government Accountability Office, 2012). Simply put: Students with disabilities must be provided reasonable testing accommodations, including Braille, sign language interpretation, large print, auditory presentation and description, and extended time on time-limited tests (Casey, 1987; Centra, 1986; Elliott, McKeivitt, B., & Kettler, 2002; Hishinuma, 1998; Phillips, 1994; Pitoniak & Royer, 2001; Thurlow, Seyfarth, Scott, & Ysseldyke, 1997; U.S. Department of Justice, 2014).

Extended time on time-limited tests is typically operationalized as “time and a half,” meaning that students provided this accommodation are allowed 50% more time to complete their tests (Laitusis, Morgan, Bridgeman, Zanna, & Stone, 2007; United States Government Accountability Office, 2012). However, most students who receive extended-time accommodations do not use all the additional 50% testing time.

For example, in one study of 825 students with learning disabilities at a small arts and technical college, very few students who were provided extended time used more than the standard time, and those who did rarely used more than an additional 25% (Holmes & Silvestri, 2019). Similarly, in another study of 605 students with disabilities at a regional university, most students who were provided extended time used only the standard time, with primarily students with psychiatric disabilities using more (around half of the additional 50%; Spenceley & Wheeler, 2016). In another study of 50 students with a variety of learning disabilities taking the SAT, no student used more than 14% extra time (Cahalan-Laitusis, King, Cline, & Bridgeman, 2006).

As Gernsbacher (2015, p. 35) wrote, “When students request extended time or time and a half, what they are really requesting is not to feel the pressure of time ticking off; not to experience anxiety about
running out of time; not to have [an untimed] power test administered as a [time-limited] test” (see also Hadley, 2007). The studies demonstrating how little additional time students with disabilities actually use support this interpretation. In essence, students with disabilities who are provided extended time are taking tests as untimed power tests, but the rest of the class is not. In this way, students with disabilities are treated differently.

Students who use extended-time accommodations are treated differently in another way as well. To implement extended-time accommodations, most universities physically segregate students with disabilities away from their peers, in separate testing locations (Liasidou, 2014). Disabled students are literally excluded from their classrooms. Constitutional Law professor Colker (2019) argues that such segregation violates the spirit, if not the letter, of U.S. disability laws. A more inclusive campus would offer all students the opportunity to experience examinations as untimed power tests in the same room.

In the testing accommodations world, accommodations are frequently distinguished from modifications (Thurlow, Lazarus, Thompson, & Morse, 2005), which is important when evaluating extended time as an accommodation. An accommodation, in contrast to a modification, must not “alter the construct measured” (Sireci, Scarpati, & Li, 2005, p. 460); must only “alter … an unimportant aspect of the test administration” (Lovett, 2014, p. 81); must only be “incidental to the construct being measured” (Brooks, 2003, p. 4); must “only change how a construct is being measured, with the goal of more valid assessment of students with disabilities” (Stretch & Osborne, 2005, p. 2); and must only “increase access to the test content while allowing for the score to be interpreted in the same manner as that for a test taken without an accommodation” (American Educational Research Association, American Psychological Association, and National Council on Measurement in Education, as cited in Cawthon & Leppo, 2013, p. 364).

Therein lies the rub. For extended time to fit the definition of an accommodation rather than a modification, it must not alter the construct measured. Put another way: Standard time-limits must, by definition, be only peripheral to the construct measured, immaterial to the test, and incidental to the assessment, in the same way that presenting a test in print rather than Braille would be. The vast timed-testing literature we
reviewed above tells us that is certainly not the case. Therefore, we recommend that time limits be removed for all students for all power tests used in educational settings.

**Time-Limited Tests Are Less Equitable**

Extended time on time-limited tests is a frequent disability accommodation (Casey, 1987; Gregg & Nelson, 2012; Lombardi, Murray, & Gerdes, 2012; Newman, Wagner, Cameto, & Knokey, 2009). In a recent audit, the United States Government Accountability Office (2012) reported that the vast majority of test-related accommodations were for extended time. Extended time on time-limited tests is also considered a “reasonable accommodation,” at least to the degree that most faculty are willing to provide the accommodation.

For example, of the 255 surveyed members of the Association on Handicapped Student Service Programs in Post-Secondary Education, 98% agreed that extended time was a reasonable accommodation; in contrast, less than 1% agreed that allowing proofreaders to replace less with more sophisticated vocabulary in students’ draft papers was a reasonable accommodation (Bumba & Goodin, 1986). Of the 420 faculty surveyed at a large public university, 74% reported being very willing to provide extended time; in contrast, only 27% reported being very willing to allow disabled students to complete alternate assessments (e.g., writing a paper, making a presentation, or doing a project in place of a time-limited test; Vogel, Leyser, Wyland, & Brulle, 1999).

Of the 107 faculty surveyed at a small college, 85% reported they would allow disabled students extended time on their exams, but only 33% reported they would allow disabled students to earn extra credit that was not offered to non-disabled students (Nelson, Dodd, & Smith, 1990). Of the 171 faculty surveyed at a research university, 97% reported giving disabled students extended time on their exams, but only 12% reported giving disabled students partial credit for otherwise wrong exam answers (Smith, 2007; see also Houck, Asselin, Troutman, & Arrington, 1992; Matthews, Anderson, & Skolnick, 1987; Norton, 1997; Skinner, 2007; Sweener, Kundert, May, & Quinn, 2002, for similar data from community colleges,
private liberal arts colleges, and other institutions).

Thus, extended time on time-limited tests is requested often and considered reasonable, at least in terms of test administrators’ willingness. But is it equitable? The premise of providing extended time is to enable students with disabilities to demonstrate a potential that would otherwise be thwarted by standard time limits. Therefore, the Maximum-Potential Thesis postulates that extended time should aid only students with disabilities (Zuriff, 2000). However, scores of studies, codified by meta-analysis (e.g., Duncan & Purcell, 2019; Gregg & Nelson, 2012; Zuriff, 2000) and systematic review (e.g., Cawthon & Leppo, 2013; Lovett, 2010; Pitoniak & Royer, 2001; Sireci et al., 2005) show that not to be the case. As we have known for a century: Many students, including those without disabilities, are “relatively inefficient in such timed … tests … [but] are able to do relatively efficient and accurate work when allowed to work more slowly” (Mudge, 1921, p. 161).

An alternative to the Maximum-Potential Thesis is the Differential-Boost Hypothesis, which postulates that students without disabilities might be aided by extending or removing time limits, but students with disabilities should be aided more (Fuchs & Fuchs, 2001). Again, numerous studies show that not to be the case. Rather, numerous studies show that removing time limits boosts the performance of numerous students, including students who are learning English (Mullane & McKelvie, 2001; Pennock-Roman & Rivera, 2011; Solórzano, 2008; Wilson, 1989; Young & King, 2008), students from under-represented backgrounds (Ardila, 2005; Knapp, 1960), and students who are older than average (Cattell, 1943; Foos & Boone, 2008). Removing time limits also attenuates stereotypic gender differences (Camarata & Woodcock, 2006; De Paola & Gioia, 2016).

Time-limited tests are also less equitable to students with disabilities who forego obtaining documentation and are, therefore, ineligible to receive extended-time accommodations. A wealth of data point to the fact that many students with less visible disabilities, such as anxiety, depression, chronic health conditions (e.g., Crohn’s disease, lupus), dyslexia, and other learning disabilities forego obtaining instructional accommodations because of stigma (Barnard-Brak, Lectenberger, & Lan, 2010; Cole & Cawthon, 2015;

The stigma of disability is particularly acute for military veterans (Kranke, Weiss, & Constanine Brown, 2017), immigrants (Nadeem et al., 2007), first-generation students (Stebleton, Soria, & Huesman, 2014), and members of cultural minority groups (e.g., Blacks: Ward, Wiltshire, Detry, & Brown, 2013; Asian Americans: Saetermoe, Scattone, & Kim, 2001; Latino men: McDonald, Keys, & Balcazar, 2007).

Some students with disabilities also forego obtaining instructional accommodations because of the exorbitant expense of obtaining the required documentation (Grimes et al., 2019; Lightner et al., 2012). As the United States Government Accountability Office (2012) reports, the expense of securing the professional evaluations necessary for disability documentation can amount to several thousand dollars. These costs are almost always borne by the student (or their family) rather than the university (Denhart, 2008; Shipp, 2008), leading to substantial economic barriers for low- and middle-income students and creating disturbing inequities in who receives instructional accommodations (Colker, 2019; Gormley, Hughes, Block, & Lendmann, 2005; Ragosta, 1987; Ragosta & Wendler, 1992; Wolanin & Steele, 2004).

As the 2019 college entrance scandal vividly illustrates, students from affluent families are considerably more likely to obtain extended-time accommodations. For example, students from affluent public high schools (where only a slim minority of students qualify for free or reduced-cost school lunches) are nearly three times more likely to receive extended-time on college entrance exams than students from low-income public schools (where the majority qualify for free or reduced cost lunches, Belkin, Levitz, & Korn, 2019). Across the nation, students whose families can afford more expensive private college tuition are more likely to have extended-time accommodations (Vickers, 2010). At state universities, students whose families can afford more expensive out-of-state tuition are more likely to have extended-time
accommodations (McGregor et al., 2012; see also California State Auditor, 2000; Griggins, 2005; Lerner, 2004).

These inequities could be attenuated by applying Lovett’s (2010, p. 616) recommendation: “if at least some nondisabled examinees would also benefit from the accommodation … it would be inappropriate to withhold the accommodation from them.” Put another way, perhaps it should be instructors who are required to justify, with documentation, why they administer time-limited tests rather than students who are required to justify, with documentation, why they need extended time (Colker, 2019), which leads to our suggestion for further research.

**Further Research**

Why, nearly a century later, do we continue to “hear school men very authoritatively saying that the fast students make the best grades and the slow ones the poorest?” and why do instructors continue to erroneously assume “that if a student knows the subject in which he is being tested it should follow that he requires but a short time to make his answer?” (Longstaff & Porter, 1928, p. 638). As instructors, we are not assessed by how speedily we can demonstrate our mastery – for instance, how quickly we can teach a semester’s worth of psychology? Why then do we continue to assess our students by how speedily they can demonstrate their mastery? Why do we continue to administer time-limited exams, which are less valid, reliable, inclusive, and equitable? These are questions that summon further research.

We have heard some instructors claim that time-limited tests are needed to ensure fluency in a subject. These instructors often point to math and second language learning as arenas for which time-limited tests are appropriate for building or assessing fluency. However, scholars of math and language learning refute these assumptions.

For example, in a commentary in *Education Week*, Stanford professor Boaler (2012) chastises teachers and administrators for falsely assuming that mathematical fluency can be developed through time-limited testing. Rather, mathematical fluency and the ability to work facilely with numbers arises from
unhurriedly acquiring conceptual knowledge not from anxiously enduring rapid-fire drilling (Boaler, 2014; 2018-2019). Many math educators question whether time-limited testing can even assess, much less develop, fluency (Clarke, Nelson, & Shanley, 2016); some directly deem it ineffectual (Kling & Bay-Williams, 2015), the source of the country’s lackluster mathematical achievement (Flynn, 2017; Pink, 2018), a ritualistic holdover from teachers’ own flawed education (McCloskey, 2014), driven by an erroneous conception of what mathematical skill acquisition is all about (Seeley, 2016). Consequently, the National Council of Teachers of Mathematics (2014) caution against the use of time-limited tests for acquiring or demonstrating mathematical fluency.

Turning to second language instruction, the debate over time-limited versus untimed tests in second language instruction dates as far back as any of the topics we have covered in this article (Carroll & Sapon, 1959, as cited in Carroll, 1990). Even in those earliest days, second language instructors questioned rote drilling as a means for developing fluency (Crawford & Leitzell, 1930, as cited in Nadal, 1950), and researchers demonstrated that fluency was independent of speed (Carroll, 1958). Contemporary educators continue those cautions: When acquiring a second language, speed does not equal proficiency (Laufer & Nation, 2001), and speeded responding should not serve as a proxy for fluency (Segalowitz, 2003) or strength of association (Hernandez-Chavez, Burt, & Dulay, 1978).

In summary, future research should investigate why instructors continue to cleave to the false beliefs that fluency is evoked and revealed by time-limited testing.

**Practice-Based Recommendations**

Because time-limited tests are less valid, less reliable, less inclusive, and less equitable, we offer the following recommendations for avoiding time-limited tests in higher educational assessment:

1) Remove all time limits from all higher educational tests intended to assess power. In addition to improving the tests’ validity, reliability, inclusivity, and equitability, removing time limits from power tests allows students to attenuate their anxiety (Faust, Ashcraft, & Fleck, 1996; Powers, 1986), increase their
creativity (Acar & Runco, 2019; Cropley, 1972), read instructions more closely (Myers, 1960), check their work more carefully (Benjamin, Cavell, & Shallenberger, 1984), and learn more thoroughly from prior testing (Chuderski, 2016).

2) If administrative constraints obviate removing time limits (e.g., a testing room cannot be reserved for an unlimited period of time), administer untimed asynchronous tests (such as take-home exams or untimed online exams). Allow students to work at their own pace and to take advantage of course materials, as well as online information.

If instructors are worried that unlimited testing time in un-proctored testing situations (take-home or online exams) will enable students to “look up the answers,” it might be the questions, rather than the students, that need examination. As advised by award-winning educator Geurin (2019), if we are worried about students accessing the Internet to complete exams, perhaps our exams are not asking the right questions: “Google doesn’t have answers. It has information. Students find answers through good thinking.” As astrophysicist Mack (2018) explains, “A surprisingly large part of having expertise in a topic is not so much knowing everything about it but learning the language and sources well enough to be extremely efficient in Google searches” (see also Boyle, 2018; Coleman, 2019; Edmiston, 2019; Rohrer, 2019). Similarly, as Gernsbacher (2014, p. 3) cautions, “if the answer to a test question is just a click away,” it might not be a very good test question.

3) Assess mastery through non-test mechanisms, such as projects, reflections, and other performative and demonstrative means of assessment. Gernsbacher’s undergraduate “Research Methods” (https://online225.psych.wisc.edu) and “Psychological Effects of the Internet” (https://internet.psych.wisc.edu) courses illustrate that a large-enrollment foundational course and an upper-division undergraduate survey course can be conducted with non-test mechanisms of assessment that better parallel the non-test mechanisms of assessment universally present in post-college life.
Footnotes

1 All data and materials supporting the conclusions drawn in this article are available in Gernsbacher, Soicher, and Becker-Blease (2019), which is a Technical Report available on Open Science Framework.

2 Although some scholars cite Thorndike (1914) as support for the assumption that faster students perform better, Thorndike’s study of college students computing simple addition problems does not provide this support. The accuracy of the fastest students did not differ from the accuracy of the slowest students.

3 If the testing room is no longer available, consider allowing students to return to the instructors’ office to continue working on their tests.

References

https://doi.org/10.1093/jpepsy/23.2.87


https://doi.org/10.1016/j.intell.2016.02.004

https://doi.org/10.1177/001316445401400308

https://doi.org/10.1007/BF02289023


Boyle, S. [stephanieboyle]. (2018, October 23). I’ve been programming for about 5 years and today googled ‘how to transpose a dataset’, because I hadn’t done it for 2 weeks and couldn’t remember … Coding isn’t about memorising; if it was, I definitely wouldn’t have a job [Tweet]. https://twitter.com/stephanieboyle/status/105481132324732934


Burack, B. (1967). Relationship between course examination scores and time taken to finish the examination, revisited. *Psychological Reports, 20, 164.* https://doi.org/10.2466/pr0.1967.20.1.164


California State Auditor (2000). *Standardized tests: Although some students may receive extra time on standardized tests that is not deserved, others may not be getting the assistance they need (2000-108).* Sacramento, CA: Bureau of State Audits.


Clarke, B., Nelson, N., & Shanley, L. (2016). Mathematics fluency—more than the weekly timed test. In K. D. Cummings & Y. Petscher (Eds.), *The fluency construct (pp. 67-89)*. New York: Springer. [https://doi.org/10.1007/978-1-4939-2803-3_3](https://doi.org/10.1007/978-1-4939-2803-3_3)


Coleman, G. [GraemeColeman]. (2019, February 20). Hello, my name is Graeme, I have a PhD in computing, and I am a senior accessibility consultant, but when I want to type “é” on a Windows laptop I go to Beyoncé’s Wikipedia page and copy/paste the letter from there [Tweet]. https://twitter.com/graemecoleman/status/1098253794691944448


Dodonova, Y. A., & Dodonov, Y. S. (2013). Faster on easy items, more accurate on difficult ones: Cogni-
tive ability and performance on a task of varying difficulty. *Intelligence, 41*, 1-10.


handbook for the Scholastic Aptitude Test and Achievement Tests* (pp. 37-67). New York: College Entrance Examination Board.


Ebo, B. A. (2016). *Understanding the experiences of college students with learning disabilities*. (Un-
published doctoral dissertation.) Northeastern University, Boston, Massachusetts.

Edmiston, E. K. [EKaleEdmiston]. (2019, January 4). A friend/colleague who is an excellent programmer offhandedly told me the other day that coding is 90% googling error messages & 10% writing code. Until this point, I thought that all the time I spent [Tweet]. https://twitter.com/ekaleedmiston/sta-
tus/1081221822186696706

Elliott, S. N., McKevitt, B. C., & Kettler, R. J. (2002). Testing accommodations research and decision mak-
ing: The case of “good” scores being highly valued but difficult to achieve for all students.


Geurin, D. [@DavidGeurin] (2019, September 9). If students are able to rely on the Internet to get answers, are we asking the right questions? Google doesn't have answers. It has information. Students find answers through good thinking. [Tweet].


Mack, K. [AstroKatie]. (2018, December 8). A surprisingly large part of having expertise in a topic is not so much knowing everything about it but learning the language and sources well enough to be extremely efficient in google searches. [Tweet]. https://twitter.com/AstroKatie/status/1071442842873159681


McDannell, L., & Lyvers Peffer, P. A. L. (2013). *The relationship between exam completion time and exam score in an introductory animal science course* (Unpublished senior honors thesis). The Ohio State University, Columbus, Ohio. [https://kb.osu.edu/handle/1811/54580](https://kb.osu.edu/handle/1811/54580)


Rohrer, B. [brohrer]. (2019, October 8). I was just asked whether I ever have to look things up when I code. I want to go on record saying that, aside from canned white board coding examples, I can’t write two lines of code without referring to Google or stack overflow. I would be lost without them. [Tweet]. https://twitter.com/brohrer/status/1049388373541089280


32


Thorndike, E. L. (1914). On the relation between speed and accuracy in addition. *Journal of Educational Psychology, 5*, 537. [https://doi.org/10.1037/h0075308](https://doi.org/10.1037/h0075308)


Minneapolis, MN: National Center on Educational Outcomes.


Vernon, P. A., & Kantor, L. (1986). Reaction time correlations with intelligence test scores obtained under either timed or untimed conditions. *Intelligence, 10*, 315-330. [https://doi.org/10.1016/0160-2896(86)90002-4](https://doi.org/10.1016/0160-2896(86)90002-4)


Research, 62, 185-194. https://doi.org/10.1097/NNR.0b013e31827bf533


