

The GED and the Rise of Contextless Accountability

ETHAN L. HUTT

University of Maryland, College Park

Background/Context: Policy discussions in the U.S. and abroad have become increasingly studded with reference to the results of international tests like PISA. Unlike most assessments, PISA is not designed to measure whether students have mastered a particular school curriculum but rather provide a measure of students' ability to meet future challenges irrespective of where in the world they live. Though growing in influence, the concept of a "contextless" form of accountability has an important antecedent in the history of American education: the Tests of General Educational Development (GED), which were developed in the 1940s to assist the transition of American World War II servicemen and women.

Purpose: The purpose of the study is to use the history of the development and subsequent spread of the GED examine the general challenge posed by contextless accountability measures.

Research Design: This study draws on a wide range of primary and secondary sources to present a historical analysis of the development and diffusion of the Tests of General Educational Development.

Conclusions/Recommendations: Noting the strong parallels between the history of the GED and the current popularity of international measures like PISA, this paper examines the history and development of the GED in order investigate the allure, promise, and pitfalls of contextless assessment and accountability. In so doing, this paper illustrates the importance of quantification as a means of creating useful abstractions but also the inherent danger of the perceived certainty of these kinds of metrics. In the decades following the 1940s, the GED retained its reputation as an objective, readily available, measure of high school achievement that could be used in any context and with any population—a task it was never intended or designed to fulfill. Thus, this paper argues, the American experience with the GED offers important lessons and insights in a world where PISA continues the reign of contextless, test-based accountability systems. Namely, that the level of abstraction required to develop these measures makes them ill-suited to inform the kind of specific policy discussion in which they are frequently invoked.

Given the ritual and pervasive handwringing about American school quality, even the most casual observer of education has come into contact with the new global dimensions of education accountability. Even if average person has never heard the acronym PISA, what it stands for, or, what, precisely, it measures, they have no doubt heard its verdict that American children are falling behind their international peers. The presence of this “global achievement gap,” as one scholar has termed it (Wagner, 2008), has become a key fact for claims that American schools are outdated, failing, and a threat to national security (e.g., Council on Foreign Relations, Klein, Rice, & Levy, 2012; Hanushek, Woessmann, & Peterson, 2012).

Though they are likely familiar with these critiques, the average American is likely less familiar with the test that produced them. Indeed, they might be surprised to learn that the impressive list of indictments against the school are generated by a test that is expressly designed not to measure how effective schools are at teaching their curriculum or how much of that curriculum students have learned. Instead, the tests are designed to assess “young people’s ability to use their knowledge and skills to meet real-life challenges, rather than on the extent to which they have mastered a specific school curriculum” (OECD, 2000, p. 14). The question is what we should make of the results of an assessment that at once proclaims our schools to be internationally inferior while at the same time professing little interest in the specific curricular tasks we have assigned to our schools.

There is already growing scholarship aimed at taking a more critical look at PISA, its constructs, its politics, and its influence even as the evidence grows that PISA results have become factors in policy debates (e.g., Duncan, 2010). Yet, the relative newness of international assessments like PISA makes it difficult to imagine what their long-term impact will be or even to gain sufficient perspective to understand all of the dimensions of their influence. This newness, however, does not prohibit drawing any conclusions about PISA and its influence. In these instances it is useful to consider historical cases that can provide perspective and insights as we begin to consider what role PISA does, and should, play in our national and international discourse on education. The early 20th century American experience with achievement testing provides a useful point of comparison with in the Tests of General Educational Development (GED).

At first blush the GED might seem like a strange choice for comparison. After all, what does a test that was developed during World War II as a tool for assessing returning veterans have to tell us about tests purporting to measure the 21st century fate of international 15 year olds? Though the

GED and PISA are different in many significant ways, their origins and use include several important parallels that suggest a consideration of the history of the GED could be instructive to those considering the future of global accountability. The most crucial of these features is their joint claim to provide a context-free measure of school outcomes. In the same way that the OECD claims PISA goes beyond the curriculum of any one school, province, or country to measure the long-term impact of school, the GED was designed to “consist primarily of those elements of knowledge which are of widest application or of greatest functional value” (American Council of Education (ACE), 1945)—what the test’s primary author described as the “lasting outcomes” of high school as opposed to “detailed descriptive facts” that serve “only as a means to these ends.” (Lindquist, 1944, p. 366). This unique orientation—something that set the GED apart from traditional achievement tests or high school entrance examinations of its time—was essential to its task: How else could the GED serve as a universal measure of high school for an American school system that was decentralized and lacking in uniformity than by eschewing content for long-term outcomes?

This claim of a context-free, geographically independent measure proved important not only to the GED’s ability to convince American lawmakers in the 1940s to accept its results, but, as we will see, it was also central to its rhetorical power and capacity for rapid diffusion. Once introduced, the GED’s influence was difficult to contain in no small part because of its claim to universality. It became a free-floating measure that could be used in any number of contexts to assess, critique, or define the legitimate outcomes of the school system. Understanding this aspect of the GED’s history is a crucial part of understanding how new school measures like PISA, which rest on similar claims of context-free assessment, have the potential to transform the idea of school even as they measure it. What Theodore Porter has claimed about the power of numbers in general can be said of these sorts of assessments in particular, they “have the power to create new things and transform the meaning of old things” (Porter, 1996, p. 19). In the case of the GED, the introduction of this new measure redefined the objectives of a high school education.

This paper proceeds by providing a brief history of the circumstances and context for the development the GED—the first universal measure of high school attainment in America. This history is followed by an examination of the rhetoric of quantification and its claim to a context-free measure provided the means for the test to break out of the context and purpose for which it was created and come to be used as a universal measure of high school quality. I conclude by offering a brief

reflection on the nature of contextless assessments like the GED and PISA, in which I argue that the level of abstraction necessary required of universal measures of this kind make them inherently difficult to control—a fact that must be acknowledged as feature of all global accountability efforts.

ONE TEST TO MEASURE THEM ALL

On January 3rd and 4th of 1942, a thousand of the nation's university presidents and leading educators gathered in Baltimore to discuss how their institutions could best support the war effort and to consider the imagined futures of the millions of men and women serving in the military. The issue that troubled the education leaders who had gathered in Baltimore was how their institutions could act quickly and decisively to devise a plan for reabsorbing and appropriately placing discharged soldiers (Tyler, 1944). The problem that these leaders glimpsed at the beginning of 1942, when the American war effort had only just begun, would grow exponentially in the coming years. By the end of the war America was faced with the challenge of reabsorbing nearly 16 million World War II veterans.

Though it had long been recognized among military and civilian leadership that America's post-secondary institutions would play a central role in this process (Altschuler & Blumin, 2009; Loss, 2012)—a feeling ultimately expressed in Title II of the GI Bill—it was not as easy as simply offering returning veterans education benefits, however generous. The difficulty, often elided in traditional accounts of the GI Bill, is captured in several basic statistics about America's returning veterans: 59% of white veterans and 83% of black veterans had not graduated from high school and a full 26% of white veterans and 55% of black veterans had never attended high school at all (Mettler, 2005, p. 56; Smith, 1947, p. 250). The limited prior academic achievement posed a serious problem to those who viewed post-secondary education as the primary mechanism for reintegrating veterans, because a high school diploma was increasingly necessary not only for college enrollment but for licensure in a growing number of occupations (cf. Council of State Governments, 1952; Rose, 1991).

These problems were mediated by two countervailing beliefs: the first was that the military had been an important educational experience for many veterans. Not only had the military gone to great lengths to train its soldiers for combat but the military had also undertaken one of the largest educational enterprises ever attempted when it created the United States Armed Forces Institute (USAFI). The USAFI represented a joint venture between the military and the University of Wisconsin to provide

(with the assistance of 85 schools, colleges, and universities) high school and college level correspondence courses to men and women serving in the American military anywhere in the world. Education was considered crucial to the mental health and morale of those serving in the military and therefore the USAFI was seen as an important part of the war effort, with more than 1.25 million servicemen and women enrolling in courses by the end of war (Loss, 2012). As an extension of the traditional privileges afforded American veterans after war (e.g., Skocpol, 1995; Skrentny, 1996, Chapter 3), many felt strongly that veterans deserved to receive “full and fair academic credit for military experience” (Charters, 1947, p. 17).

The second, and related, belief was that for the first time the technical psychometric tools existed to allow for the scientific measurement and accreditation of veterans’ educational experiences. After the World War I, post-secondary institutions had felt similarly about both the need to value military service and the educational value of that service but there were no available tools for the task. Instead, colleges had simply offered veterans “blanket credit”—credit awarded in relation to the amount of time served (American Council on Education, 1943). Educators feared that, given the sheer size of the demobilization effort, if this practice were repeated, it would severely damage the value of academic diplomas. Rather than issue credit on time served, they reasoned, credit should be issued on the “basis [of] competence actually demonstrated through performance on specifically prepared examinations” (American Council on Education, 1943).

The questions that remained was what form these examinations should take and whether they could produce an assessment that addressed the most pressing needs of educators and veterans. Given the view reflected in the GI Bill that the opportunity to continue their education was central to the veterans’ readjustment and long-term job prospects as well as the belief that sending war veterans back to high school to complete their degrees would be degrading, the issue of assessing veterans’ education levels was framed in terms of high school equivalency. The question of how to devise a test to measure whether veterans, through their experience in the military, had attained the equivalent of a high school diploma, fell to the renowned University of Iowa testing expert E. F. Lindquist.

Today Lindquist is perhaps most famous for his series of accomplishments in the 1950s and 1960s when he revolutionized the grading of standardized tests through the introduction of Optical Mark Reader (OMR) technology; created the American College Testing program (ACT) as a rival to the Educational Testing Services’ SAT; and founded

the Measurement Research Center (Feldt, 1979), which would later be purchased by Westinghouse Learning Corporation and subsequently National Computer Systems and Pearson. The level ambitious and creativity that would come to mark Lindquist's involvement in these later projects was already evident in his early work as a psychometrician. Lindquist was widely known for his work in developing the Iowa Test of Educational Development (ITED), which was used in the state and around the country to conduct scholarship competitions for high school students seeking to go to college (Peterson, 1983). Lindquist was also the author of the widely read and influential textbooks *A First Course in Statistics* (1938), *Sampling in Educational Research* (1940), and *Statistical Analysis in Educational Research* (1940), which were used to train several generations of education researchers in how to utilize statistical analysis techniques in their work.

The belief that a standardized test would be the solution to the problem posed by demobilization along with the decision to tap a scholar like Lindquist to solve it reflected both the considerable faith of psychometricians in the power of educational testing but also of the considerable prestige that the field of educational measurement had achieved by the end of the war. Prior to World War I, educational testing was a small field given little respect or attention inside or outside of the field of education. That changed with the creation of Army Alpha and Beta tests during the WWI, the introduction and the spread of "IQ Testing" during the interwar years, and the general proliferation of mental testing and personnel management techniques in business (Carson, 2007; Gould, 1996; Kett, 2013). The outbreak of World War II did little to abate the rise of educational testing as the army turned to the Army General Classification Test as a means of sorting newly enlisted soldiers into the positions where they would be most useful.

Lindquist's initial proposal was to adapt ITED, which had been designed to measure students' broad educational outcomes regardless of the specific school they attended, to fit the current and somewhat analogous situation. The difficulty was that while ITED had been designed to account for possible variation within Iowa high schools, the test Lindquist would produce for veterans had to account for learning that had not taken place in high school at all. As Lindquist explained in his definitive statement on the philosophy and approach of the Tests of General Educational Development (GED), the GED was "designed especially to provide a measure of a general educational development which results from . . . all of the possibilities for informal self-education which military service involves, as well as the general educational growth incidental to military training an experience as such" (Lindquist, 1944,

p. 364). The kind of informal experiences Lindquist had in mind ranged from the exposure to foreign languages, job experiences, social customs, and physical and economic geography of the places soldiers had been stationed during the war to the “reading of newspapers, magazines and books, self-directed study and deliberation, educational movies, lectures, formal and informal discussions, correspondence with friends at home, etc.” (Lindquist, 1944, p. 359). In order to measure these disparate, informal activities Lindquist and his colleagues devised a battery of five tests: Correctness and Effectiveness of Expression; Interpretation of Reading Materials in the Social Studies; Interpretation of Reading Materials in the Natural Sciences; Interpretation of Literary Materials; and General Mathematical Ability (American Council on Education, 1945). These would be the five-test battery comprising the GED. In its final form the test would take 10 hours to administer.

Distilling the entirety of a high school education into a single test battery, the results of which could secure a high school equivalency diploma for the test taker, ultimately required Lindquist to decide what elements were at the core of a high school education. In a section entitled “Desirable General Characteristics of the Tests,” the *Examiner’s Manual* makes clear the stance the test’s creators have taken on the question of which outcomes of a high school education need to be measured in order to establish equivalency. The manual explains, “the tests must measure as directly as possible the attainment of the *ultimate* objectives of the whole program of general education, and must minimize as much as possible the more immediate and temporary content of special school subjects” (American Council on Education, 1945, p. 6). These ultimate objectives, according to the test’s official documentation, were “those elements of knowledge which are of wide applicability or of greatest functional value” (American Council on Education, 1945, p. 6).

Embracing this view of the meaning and value of a high school education, of course, allowed Lindquist to deal with the most difficult aspect of his challenge: coming up with a test that could be given to veterans regardless of where or if they were formally educated in school. Minimizing the specific content maximized the number of people it could claim to assess fairly. Lindquist and his team had meet the specific needs of America’s demobilization challenge, they had produced a universal measure of high school equivalency—one that could function anywhere and on anyone.

Though the decision to minimize the academic content of the GED well served the demands of the time, it must be noted just how radical a vision of high school the GED represented. In his writings, Lindquist justified the lack of focus on curricular content by saying, “It is generally recognized that the lasting outcomes of a high school . . . course are

not the detailed descriptive facts which are taught—most of these are forgotten by the typical student within a short time after he completes the course.” Rather these “detailed materials of instruction” are “to a large extent only a means toward [the] ends” of a high school education. (Lindquist, 1944, p. 366). They believed that academic content knowledge could be so sufficiently separated from the lasting outcomes of a high school education that Lindquist and his colleagues expressly designed the test to prevent strong correlations between subject achievement tests and the GED. After all, high correlations between the achievement tests and the GED, Lindquist reasoned, would indicate an overreliance on formal school skills that would place the veteran, who might not have attended high school, at a disadvantage (Lindquist, 1944; cf. Mosel, 1954).

Despite the GED’s strong commitment to going beyond the content of a high school curriculum, this view was far from being the “generally recognized” idea that Lindquist suggests. In fact, the extent to which the subject matter of the high school curriculum was meaningful as a result of, or in spite of, its specific content was a matter of considerable and protracted debate among educators (e.g. Kliebard, 2004). The lack of agreement or uniformity on this issue is evident in state laws concerning the courses necessary to earn a high school diploma—the standard the GED was intended to replicate. For instance, high school diploma requirements for the state of New York required all students to take four units of English, one unit of science, three units of social studies, and one unit of health, but the state had no minimum math requirement. Iowa by contrast required three units of English, three units of social studies, one unit in physiology and hygiene, 3.5 units of math, but had no minimum science requirement, while the state of California required only that all students take a civics course and left the specific high school diploma requirements to individual districts (Hess, 1946, pp. 126, 108–109, 99).

The lack of agreement on the course of study that would produce the lasting outcomes of high school aside, the GED was equally radical in its basic premise that the ultimate purpose of high school was a set of learning outcomes. This view was in sharp contrast to the longstanding American legal interpretation of compulsory school laws (Hutt, 2012). Interpreting those laws, judges had expressly rejected the idea of education embodied in the GED—that the purpose of education could be reduced to discrete knowledge outcomes. The educational experience—the one that justified compelling all children to attend school—was something that was intimately tied to the context of the school. This was, in part, why judges had considered the age requirements in compulsory school laws to be

minimum time requirements rather than maximum time requirements. That is, even if a student had reached the state’s minimum learning requirements, the child still had to attend school because the purpose of school attendance was not just about acquiring knowledge.

One particularly revealing illustration of the extent to which the GED represented an extreme narrowing of the purpose of high school, the 1945 *Annual Report* of the Maryland State Department of Education, included a fold out chart of the individual student and his relationship to the curriculum. There were 15 items listed under the heading “The Ends for Which We Educate” on the far right of the chart. The aims listed represented a broad and holistic vision of education ranging from the moral “reverence for ad practice of *sound ethical and moral principles*” to the aesthetic “Growing appreciation of living through *literature, art, music*”; to the social “realization of the importance of and practice in experiencing satisfying *human relationships*”; to the conservational: “a respect for the worth of the individual and an awareness of the importance of the physical environment, resulting in a sound attitude toward the *conservation of human and natural resources*” (Maryland State Board of Education, 1945, p. 189, emphasis in original). Of all the aims listed, only one of them—“competence to the level of one’s ability in reading, writing, listening, computing, and thinking”—could plausibly be construed as measured by the GED and only then depending on one’s definition of “thinking” and “to the level of one’s ability” (Maryland State Board of Education, 1945, p. 189). There could be little doubt that the GED represented a radical redefinition of the stated purpose of a high school education and that this definition had been developed by the very small group of people worked for the USAFI and ACE.

To recognize the results of the GED as the basis for issuing an equivalency diploma suggested that, whatever else a high school education might entail, it was the ability to demonstrate certain learning outcomes—regardless of where that knowledge had been obtained—that made one eligible for a diploma.

And recognize it they did. By the end of the 1946, 44 of the 48 states were issuing degrees to returning veterans on the basis of GED results (American Council on Education, 1946). Of the other four states—Maine, New York, New Jersey, and Massachusetts—two of them would have policies in place by the end of the following year (Maine and New York) and the other two acknowledged the value of the GED for placement purposes but had chosen not to issue a high school diploma or equivalency certificate without additional course work (American Council on Education, 1946). Given the decentralization of the American education system and the considerable diversity in education policies among states,

the near universal adoption of the GED—and with it the embrace of the radical idea of issuing degrees by examination—was a remarkable result.

The rapid and near universal adoption of the GED as an legitimate measure of high school equivalency was a testament to the hard work of the American Council of Education and its Committee on the Accreditation of Service Experience (CASE), which had been charged with disseminating information about the GED and securing its use and legal recognition by states. CASE officials had been successful at convincing school and state officials that the GED was a solution not just to the looming threat of mass demobilization but also a means to inject rigor into modern educational standards. CASE officials argued that the state of the art testing techniques used by the GED creators, had produced a nationally normed, standardized test that represented not just an equivalent measure of high school but a superior measure as well.

For example, Cornelius Turner, the assistant director of CASE, frequently argued that the rigors of objective measurement meant that it was time to reject the common argument that “an intangible something is obtained through school attendance which a regular high school diploma represents but which can never be measured by tests” (Turner, 1949, p. 389). Indeed, the old way of doing things that involved subjective assessment by individual high schools all around the country had resulted in a proliferation of standards. “There are more than 20,000 secondary schools in this country, and for each school the diploma represents different things,” Turner explained. “This variation could be multiplied by the number of variations in the curriculum in all the schools. Each high school diploma means something different; perhaps it is time we had a secondary school credential which has some uniformity of meaning” (Turner, 1949, p. 389). Not only would the GED bring greater certainty and standardization to the meaning of the diploma but it would bring more rigor as well. In advocating for the GED’s adoption, CASE officials pointed to the fact that based on their national norming study, some 10% of graduating high school seniors would not been able to reach the recommended cut score for issuing a GED high school equivalency diploma (American Council on Education, 1945). Thus, the GED promised, finally, to provide rigor and uniformity in high school assessment—to create an independent, universal standard for the high school diploma. The rapid adoption of the test suggested that states believed in that promise.

The Imperialism of the GED

All that the GED represented might have been nothing more than an interesting footnote in the history of the post-war period if the test had remained, as initially planned and envisioned, a tool strictly for assessing the educational attainment of demobilized veterans. Instead, the test, and its equivalency credential, became ubiquitous. Despite its unilateral reimagining or radical narrowing of the meaning and content of a high school education, its lack of claim to any particular setting or curriculum allowed it to offer a readymade solution to any number of problems. By the end of the 1960s more than a quarter-million Americans would take the GED each year and by 2009, one in every seven secondary school credentials would be issued on the basis of the GED (National Center for Education Statistics, 2011). That the GED would long outlive its initial purpose is a testament to the rhetorical power and political usefulness of the kind of contextless assessment that the GED represented.

No sooner had GED been uniformly embraced by states as the solution to the issue of granting soldiers credit for their military experience than states began seeing the GED as the solution to other problems as well. If the science of educational measurement, accompanied by a state seal of approval, had deemed the test battery good enough to measure out of school learning in one context, the logic ran, why could it not perform the same function in a different context? It was a logical extension of the original premise for the test—one that was easy to make. Rather than seeing the GED as a particularistic solution, calibrated to the specific conditions of military experience, many commentators, educators, and state officials began to view the GED as a universal solution to the issue of educational measurement.

In 1946, not long after New York became the 45th state to use the GED as the basis for granting an equivalency diploma, the *Journal of Higher Education* ran an editorial praising the principles of measurement embodied in the test. Drawing a comparison between the “save your coupons and get an education” unit credit system in America and a European system overly reliant on comprehensive examinations that “makes too much turn on a single show of strength,” the editorial board considered the philosophy embodied in the GED to be a good middle ground between these extremes. Because the tests were “constructed without reference to the content of any particular course and designed to test progress in general education” the tests “could be used in connection with almost any type of curriculum organization,” the editorial board reasoned (*Journal of Higher Education*, 1946). This was a significant advance, the board argued, and one that deserved to be more broadly embraced by the American education community.

The editorial board was not alone in recognizing the broad potential of the GED tests. Noted educational researchers Hermann Remmers and Nathaniel Gage explained in an article entitled “Reshaping Educational Policy” that “educational policy has further been affected by the development of the General Educational Development Tests, which have provided a means for substituting measured achievement for credit hours or units measured by the clock and the calendar” (Remmers & Gage, 1949). Though not endorsing the use of the tests, Remmers and Gage acknowledged the increasingly widespread belief that the GED tests could and should be used in a broader context.

The major obstacle preventing its widespread use was the initial recommendations from CASE about the use of the tests. In an effort to assuage educators who objected that the GED would encourage students to drop out of school, CASE recommended that the GED “tests NOT be administered or recognized as a measure of high school equivalence until after the class of which the man was a member has been graduated” (Committee on the Accreditation of Service Experience, 1946, emphasis in original). This recommendation did not prevent states from adopting laws that specified any use for the GED that they saw fit. In February 1948 the New York State Senate and Assembly passed a law officially recognizing the GED as a means for measuring the experiences of *all adult citizens of the state* and as the basis for issuing an equivalency diploma with the same legal standing as the state’s traditional Regent’s Diploma (NY SL New York State Assembly, No. 2044, Int. 1942; Feb 12, 1948; New York State Senate, No. 1539, Int. 1471).

This was the opening gambit of what would become a nationwide trend. By April 1949, 25 states, more than half of those using the GED, had authorized the test’s use for all adult residents of the state. This number would continue to climb moving to 27 states in 1951, 31 states by 1954, and 44 states by 1963 (CASE, 1951, 1956). By 1959 the number of adults taking the GED would surpass for good the number of veterans sitting for the battery (Veterans Testing Service, 1959). Not only were more and more adults being given a chance to earn their high school diplomas via the GED, but state’s minimum age requirements set for GED test takers would fall steadily. In 1960, 46 states required individuals to be at least 20 years of age in order to receive an equivalency certificate. By 1972 that number had shrunk to just four states, while 27 states required individuals be 19 and 24 states required students be only 18. For those who wanted access to the examination and its test results without a diploma (for use in getting a job), 37 states authorized 17 year olds to take the test and five gave access to 16 year olds (CASE, 1946, 1960, 1972). The logic of these policies changes were the same as those that motivated the expansion of

the GED's use from veterans to all adults: If the test represented an objective measure of high school attainment, then age considerations were irrelevant—the test measured what it measured.

It was not just an issue of who had access to the GED that changed in the decades following the GED's introduction but the *purposes* that the test served changed as well. The use of GED scores as a condition of employment by businesses in states that retained age restrictions on equivalency diplomas also proliferated at this time (Tyler & Institute United States Armed Forces, 1956). And, as states steadily lowered the GED's age requirements to include high school age students, lawmakers increasingly turned to the GED as a readily available, legally and publicly approved measure of high school. Consistent with its long history of credentialing individuals in preparation for the labor market, the GED found its way into federal legislation like such as the Pre-discharge Education Program (PREP) for members of the military and Title I of the Economic Opportunity Act (P.L. 88-452), part of President Johnson's War on Poverty, in which the GED was considered one successful outcome of Job Corps participation (see, for example, Weeks, 1967; Weir, 1988). The GED had become the educational equivalent of nylon—a synthetic wartime substitute turned mainstream product.

The more widely used the GED became and the more people used the GED to join the ranks of high school graduates, the more the idea of the high school graduate became intertwined with the GED itself. Despite an acknowledgment that the GED was a narrow conception of high school attainment, scholars like Benjamin Bloom drew on the results of the 1955 national GED renorming study to offer the first national study of the “relationships between what is ‘put into’ the educational system and the ‘outcomes’ of the educational system” (Bloom, 1956; see also Bloom & Statler, 1957). The results of the norming study were also taken up by critics of public schools and used as evidence of their declining quality. The discovery that students in California scored lower on the 1955 norming study than they had in 1943 led to a series of articles in the *Atlantic Monthly*, in which the GED test scores were offered up as proof of the California public school system's “utter contempt for intellectual standards” (Smith, 1958). Treating the GED as not only legally equivalent but as synonymous with a high school diploma not only distorted perceptions and policy conversations around schools in the 1950s. A growing number of scholars claim that the counting of GED recipients as high graduates in official statistics has long distorted statistics on high school graduation rates (e.g., Heckman & LaFontaine, 2010; Swanson & Chaplin, 2003). As late as 2007, New Jersey allowed nonresidents to submit their GED scores and

receive a state-issued high school diploma. The state counted those people in its official graduation statistics such that in some years the state reported issuing more diplomas than students enrolled in the 12th grade (Heckman & LaFontaine, 2010, p. 247).

Taken together, these events make clear the imperialism of a contextless measure like the GED. The qualities that made it the perfect solution for the particular problem of assessing the vast and varied experiences of war veterans were the same ones that made it difficult to restrict its use to only those circumstances. The easy analogies of equivalence and the power of objective measurement made the GED appear as a solution to any number of tasks—a universal tool for accrediting disparate learning experiences wherever they were acquired in terms of diplomas, a national measure of school quality, a solution to labor market readiness. Though it found ready acceptance in all of these domains, the GED still represented a radical departure from previous efforts to measure school attainment. Out of a need to create a context-free measure of high school attainment, the creators of the GED proposed to measure not the “descriptive facts and technical data” of high school but the “ultimate objectives” the “lasting outcomes” of a high school education. In offering up the GED as a measure of “high school equivalency” Lindquist and his colleagues may have been providing an objective measure, but it was not a neutral one. The GED represented a profound redefinition of the outcomes of high school.

The Qualities of Contextless Assessments

The foregoing was a brief history of the creation, adoption, and spread of the GED as a measure of high school equivalency. For contemporary observers of education, it will also appear as a story with some remarkably strong resemblances to the recent history of PISA. Like the GED, PISA was created by a relatively small nongovernmental organization—the OECD—that sought to provide with its test a universal measure of high school quality. The authors of PISA, as with the GED, achieved this universality—this contextless quality—by abstracting from the actual curricular of individual schools, states, and countries. While the GED claimed to measure the “ultimate outcomes of a high school education” those concepts with the “greatest functional value” (American Council on Education, 1945, p. 6), PISA, likewise, claims to be “forward-looking, focusing on young people’s ability to use their knowledge and skills to meet real-life challenges” (OECD, 2001, p. 16). More specifically, the tests embody a particular worldview and a radically reduced vision of the meaning and purpose of high school. As Tröhler has argued, PISA

reflects the specific culture of Cold War efforts to “harmonize the educational globe” and the desire to place human capital production concerns at the heart of the educational system (2010). In the same way that the GED revised decades of legal and cultural understanding of education as being inseparable from the school context, PISA represents a break from the educational traditions of many its participant countries—making its constructs incoherent when interpreted in light of specific cultural and national traditions (e.g., Tröhler, 2011).

Despite these many limitations, there can be little doubt that PISA results drive headlines or that PISA results become key statistics in any number of national narratives constructed around school failure (e.g., Peterson, Lastra-Anadon, Hanushek, & Woessmann, 2011) or success in the case of Canada (e.g., Simpson, 2010) or Finland (e.g., Kupiainen, Hautamaki, & Karjalainen, 2009). Moreover, there is growing evidence of PISA’s legitimacy among high-level policy makers throughout Europe and of the increased use of PISA data to justify or motivate policy decisions (Gerk, 2009). This, despite persistent concerns of the narrowness of PISA’s strictly instrumental conception of education and the resulting sacrifice of larger questions about the multiple purposes of education (Biesta, 2009). All of these facts suggest that PISA is headed for a career similar to that of the GED in which it acts as a free-floating standard ready to be incorporated into the plans and critiques of any number of partisans, professors, or policy makers.

Though there are many striking similarities between the development and proliferation of the GED and PISA, given the different purposes of the GED and PISA, it is difficult to draw any explicit “historical lessons” from the parallel. The GED, after all, was tied up in the issuing of a specific credential and was, theoretically at least, pegged to the American high school diploma. PISA, on the other hand, remains a measure of the OECD’s best guess about the skills that will make the average global citizen successful in her future career. Even still, it may be possible to step back and say something about the nature of contextless accountability measures in general.

In *When Formality Works*, Stinchcombe argues that there are three features that allow a formalism to be successful: “cognitive adequacy,” “communicability,” and “trajectory of improvement” (2001, p. 19). As long as all three of these elements are a part of the formalism—an abstraction designed to communicate some part of an underlying phenomenon—then the formalism can function effectively. Consider, for instance, a teacher in the process of writing an algebra final. In crafting the final, the teacher will no doubt begin by consulting the statewide standards that outline the core skills and knowledge a student must possess in

order to be considered proficient in algebra. Assuming that the standards are high quality and, thus, offer a good representation of the subject, we would say the standards have a high degree of cognitive adequacy. If the teacher uses these standards to produce a well-written well-aligned final, it too will be considered to have a high degree of cognitive adequacy. The cognitive adequacy of both the standards and the test based on those standards are crucial to ensuring that the student's score on the final provides the basis for the second two elements of Stinchcombe's test. A well-designed test means that the score provides a strong indication of how well the student has mastered the underlying skills (communicability). In the case of a student who does perform well, the overall scores and sub-scores indicate places where a student needs to improve his skills. This feedback provides a trajectory of improvement for the student. In subsequent years, the scores also provide a trajectory of improvement—a feedback loop—for the standards and tests as well. If previously high-performing students appear in the next level unprepared, it offers a strong signal that the students' grades have ceased to be effective formalisms and that the standards and/or the tests need to be revisited.

This is, of course, a best case scenario for how tests could produce a useful formalism. In many cases, tests, and the standards they are designed to reflect, fail to meet these requirements, and thus the test results cease to be a successful formalism. Indeed, it is difficult to imagine how a contextless measure like the GED or PISA could ever fulfill Stinchcombe's criteria. Though the scores generated by the GED and PISA are clearly effective forms of communication, the adequacy of these forward-looking measures—of an imagined future in which the lasting outcomes of high school are still present or the skills necessary to succeed anywhere in the world in a volatile labor market—are, inherently, nearly impossible to assess. The underlying constructs are too vague to assess the robustness of the formalisms the tests produce. Moreover, there are few possibilities for the creation of a productive feedback loop—trajectory of improvement—to bring the measure more in line with the desired outcome: The future is always a moving target and one can only define success in retrospect. While acknowledging that this critique might be leveled at any attempt to quantify school outcomes, it must be said that the problem is particularly acute with contextless measures like the GED or PISA. Thus, the GED and PISA, because of their desire to measure an outcome that can never really be validated, are almost certain to become a pernicious formalism. Without a fixed point to orient toward, efforts to iterate policy in the direction of these formalisms only results in directing schools further

and further off course. Of course, recognizing this fact will do little to diminish the appeal of these types of measures—certainly not in an era when quantification and “objective” measurement have become the coin of the global realm of accountability. But it is important to point out when something is a feature of these reform and not a defect. A contextless assessment should never be the guiding star of an education system. It is hard enough in education to chart a true course let alone one toward illusory ends.

REFERENCES

- Altschuler, G. C., & Blumin, S. M. (2009). *The GI Bill: A new deal for veterans*. Oxford: Oxford University Press.
- American Council on Education. (1943). *Sound credit*. Washington, DC: American Council on Education.
- American Council on Education. (1945). *Tests of General Educational Development (high school level) examiner's manual*. Washington, DC: American Council on Education.
- American Council on Education. (1946). *Accreditation policies of state departments of education for the evaluation of service experiences and USAFI examinations*. Washington, DC: The Commission on the Accreditation of Service Experience..
- Biesta, G. (2009). Good education in an age of measurement: On the need to reconnect with the question of purpose in education. *Educational Assessment, Evaluation and Accountability*, 21(1), 33-46.
- Bloom, B. (1956). 1955 normative study of the Tests of General Educational Development. *The School Review*, 64(3), 110-124.
- Bloom, B., & Statler, C. (1957). Changes in the states on the Tests of General Educational Development from 1943 to 1955. *The School Review*, 65(2), 204-221.
- Carson, J. (2007). *The measure of merit: Talents, intelligence, and inequality in the French and American republics, 1750-1940*. Princeton: Princeton University Press.
- Commission on Accreditation of Service Education. (1946). *Accreditation policies of state departments of education for the evaluation of service experiences and USAFI examinations*. Washington, DC: American Council on Education
- Commission on Accreditation of Service Education. (1951). *Accreditation policies of state departments of education for the evaluation of service experiences and USAFI examinations*. Washington, DC: American Council on Education
- Commission on Accreditation of Service Education. (1956). *Accreditation policies of state departments of education for the evaluation of service experiences and USAFI examinations*. Washington, DC: American Council on Education
- Commission on Accreditation of Service Education. (1960). *Accreditation policies of state departments of education for the evaluation of service experiences and USAFI examinations*. Washington, DC: American Council on Education
- Commission on Accreditation of Service Education. (1972). *Accreditation policies of state departments of education for the evaluation of service experiences and USAFI examinations*. Washington, DC: American Council on Education
- Charters, W. W. (1947). Techniques of giving and taking advice: USAFI's Advisory Committee. *Educational Record*, 28, 5-20.
- Council of State Governments. (1952). *Occupational licensing legislation in the states*. Chicago: Council of State Governments.
- Council on Foreign Relations, Klein, J. I., Rice, C., & Levy, J. (2012). *U. S. education reform and national security*. New York: Council on Foreign Relations, Independent Task Force.
- Duncan, A. (2010). *Secretary Arne Duncan's remarks at OECD's release of the Program for International Student Assessment (PISA) 2009 Results* [Department of Education Press release]. Retrieved from <http://www.ed.gov/news/speeches/secretary-arne-duncans-remarks-oecd-release-program-international-student-assessment>
- Feldt, L. (1979). Everett F. Lindquist 1901-1978 a retrospective review of his contributions to educational research. *Journal of Educational and Behavioral Statistics*, 4(1), 4-13.
- Gerk, S. (2009). Governing by numbers: The PISA "effect" in Europe. *Journal of Education Policy*, 24(1), 23-37.

- Gould, S. J. (1996). *The mismeasure of man: Revised and expanded*. New York: W.W. Norton & Company.
- Hanushek, E., Woessmann, L., & Peterson, P. (2012) Is the U.S. catching up? *Education Next*, 12. Retrieved from <http://educationnext.org/is-the-us-catching-up/>
- Heckman, J. J., & LaFontaine, P. A. (2010). The American high school graduation rate: Trends and levels. *The Review of Economics and Statistics*, 92(2), 244–262.
- Hess, W. (1946). State requirements for a high school diploma for the veteran. *NASSP Bulletin*, 30, 92–144.
- Hutt, E. L., (2012). Formalism over function: Compulsion, courts, and the rise of educational formalism in America, 1870–1930. *Teachers College Record*, 114(1), 1-27.
- Journal of Higher Education. (1946). Editorial comments. *The Journal of Higher Education*, 17, 331–338.
- Kett, J. F. (2013). *Merit: The history of a founding ideal from the American revolution to the twenty-first century*. Ithaca: Cornell University Press.
- Kliebard, H. M. (2004). *The struggle for the American curriculum, 1893-1958*. New York: Routledge & Kegan Paul.
- Kupiainen, S., Hautamäki, J., & Karjalainen, T. (2009). *The Finnish education system and Pisa*. Helsinki: Ministry of Education.
- Lindquist E. F. (1940). *Statistical analysis in educational research*. Boston: Houghton Mifflin
- Lindquist E. F. (1944). The use of tests in the accreditation of military experience and in the educational placement of war veterans. *Educational Record*, 25, 357–376
- Loss, C. (2012). *Between citizens and the state: The politics of American higher education in the 20th Century*. Princeton: Princeton University Press.
- Maryland State Board of Education. (1945). *Seventy Ninth Annual Report State Board of Education of Maryland, 1945*. Baltimore: Maryland State Board of Education.
- Mettler, S. (2005). *Soldiers to citizens: The G.I. Bill and the making of the greatest generation*. Oxford: Oxford University Press.
- Mosel, J. (1954). The General Educational Development of Tests (high school level) as a predictor of educational level and mental ability. *The Journal of Educational Research*, 48(2), 133.
- National Center for Education Statistics. (2011). *Digest of Education Statistics, 2011*. Washington, DC: National Center for Education Statistics.
- OECD. (2001). *Knowledge and skills for life. First results from PISA 2000*. Paris: OECD Publishing.
- Peterson, J. J. (1983). *The Iowa testing programs: The first fifty years*. University of Iowa Press.
- Peterson, P., Lastra-Anadon, C., Hanushek, E., & Woessmann, L. (2011). Are U.S. students ready to compete? *Education Next*, 11(4). Retrieved from <http://educationnext.org/are-u-s-students-ready-to-compete/>
- Porter, T. (1996). *Trust in numbers: The pursuit of objectivity in science and public life*. Princeton, NJ: Princeton University Press.
- Remmers, H., & Gage, N. (1949). Reshaping educational policy. *Review of Educational Research*, 19(1), 77–89
- Rose, A. (1991) Preparing for veterans: Higher education and the efforts to accredit the learning of World War II servicemen and women. *Adult Education Quarterly*, 42(1), 30–45.
- Skocpol, T. (1995). *Protecting soldiers and mothers: The political origins of social policy in the United States*. Cambridge, MA: Belknap Press
- Skrentny, J. (1996). *The ironies of Affirmative Action: Politics, culture, and justice in America*. Chicago: University of Chicago Press
- Simpson, J. (2010). Canada is not becoming outclassed. *The Globe and Mail*. Retrieved from <http://www.theglobeandmail.com/commentary/canada-is-not-becoming-outclassed/article4082381/>

- Smith, M. (1947). Populational characteristics of American servicemen in World War II. *The Scientific Monthly*, 65(3), 246–252.
- Smith, M. (1958). How to teach the California child: Notes from the Never-Never Land. *Atlantic Monthly*, 202, 32–36.
- Stinchcombe, A. L. (2001). *When formality works: Authority and abstraction in law and organizations*. Chicago: University of Chicago Press.
- Swanson, C., & Chaplin, D. (2003). *Counting high school graduates when graduates count: Measuring graduation rates under the high stakes of NCLB*. Washington, DC: The Urban Institute.
- Tröhler, D. (2010). Harmonizing the educational globe: world polity, cultural features, and the challenges of educational research. *Studies in Philosophy and Education*, 29(1), 7–29.
- Tröhler, D. (2011). Concepts, cultures, and comparisons: PISA and the double German discontentment. In D. Tröhler (Ed.), *Languages of education: Protestant legacies, national identities, and global aspirations* (pp. 194–207). London: Routledge.
- Tyler, R. (1944). Sound credit for military experience. *Annals of American Academy of Political and Social Science*, 231, 58–64
- Tyler, R., & Institute United States Armed Forces. (1956). *Conclusions and recommendations on a study of the General Educational Development Testing Program*. Washington, DC: American Council on Education.
- Turner, C. P. (1949). Accreditation by means of tests. In E. G. Williamson (Ed.), *Trends in student personnel work*. Minneapolis: University of Minnesota Press
- Veterans' Testing Service. (1959). 1959 annual report of the Veterans' Testing Service, High School Level GED Testing. Washington, D.C.: American Council on Education.
- Wagner, T. (2008). *The global achievement gap: Why even our best schools don't teach the new survival skills our children need—And what we can do about it*. New York: Basic Books.
- Weeks, C. (1967). *Job Corps; dollars and dropouts*. Boston: Little, Brown.
- Weir, M. (1988). The federal government and unemployment: The frustration of policy innovation from the New Deal to the Great Society. In M. Weir, A. S. Orloff, & T. Skocpol (Eds.), *The politics of social policy in the United States* (pp. 149–190). Princeton, NJ: Princeton University Press.

ETHAN L. HUTT is an Assistant Professor in the Department of Teaching and Learning, Policy and Leadership at the University of Maryland. His research focuses on the historical relationship between schools, the law, and education policy. In particular, it examines the way in which the law has shaped public education in America through the creation of standards and the use of quantification.