**Teacher, Tester, Soldier, Spy:**

**Psychologists talk about teachers in the intelligence-testing movement,**

**1910s-1930s**

This article explores the complex rhetoric deployed by some psychologists in relation to teachers within early 20th-century research on intelligence-testing in America. Psychologists’ ways of talking to, at, and about teachers at that time set up their relationship as characterised by an originary indebtedness – of teachers towards psychology. That framework eclipsed in discourse what was, in reality, a more chaotic relationship of mutual usefulness, defiance, and interdependence between the two professions. Instead, it established between teachers and psychologists the assumption of an unrepaid – and, arguably, unrepayable – moral obligation.

Anyone who studies teachers in the Progressive Era in relation to the rise of intelligence-testing is tempted to bemoan the confiscation or ‘colonization’ of teachers’ expertise by scientists in that time.[[1]](#footnote-1) Yet, importantly, teachers were, in the minds of psychologists, the foot soldiers of the intelligence-testing movement. As Paul Davis Chapman puts it, ‘the key to the whole testing program was the classroom teacher’[[2]](#footnote-2). For intelligence-testing to spread, psychologists wanted to secure the support of teachers in buying and administering the tests; they had to not just teach teachers to test, but persuade them of what they saw as the uncontroversial rightness, benevolence and importance of testing. At the same time, psychologists could not spell their own obsolescence by granting teachers complete control of their measuring instruments.

These are to some extent familiar tensions. JoAnne Brown, in a seminal study on the semantics of profession-building in the intelligence-testing movement, theorized the pull between ‘popularization’ and ‘monopoly’ of I.Q. in its early days.[[3]](#footnote-3) Psychology, a baby discipline in the 1900s, had to self-advertise by giving information about its methods, while keeping safe its claims to expertise. Brown argues that psychologists solved this tension by using engineering and medical metaphors, which gave the public adequately evocative knowledge, but kept the actual science arcane. As far as the world of education was concerned, as Brown notes, psychologists engaged in denigrating and ridiculing teachers. While this is, as we shall see, quite true, it was not the whole story; psychologists’ rhetoric about and towards teachers was not just a show of muscles, hard facts and scientific knowledge. Intelligence-testing proponents were not (or not just) engaged in ruthless territorial conquest by wielding the weapons of statistics and (pseudo-)science. [[4]](#footnote-4) The teacher instead became for psychologists an ideal character and interlocutor through whom to operate crucial conversions from the abstract, ‘cold’, scientific object of I.Q. to a social and moral project, embodied by individuals, anchored in practice, and traversed with hope. In that story, I.Q. testing became a boon, with teachers as its primary recipients, who owed it to society and to themselves to make the most of it.

# From conquest to diplomacy: Establishing the relationship

The golden decade of intelligence-testing, between the mid-1910s and the mid-1920s, was an era well-known for the scientific scrutiny surrounding education, and the difficulty for schools to cope with a huge rise in urban population and an increasingly diverse student body.[[5]](#footnote-5) At the time, ‘virtually every aspect of education’[[6]](#footnote-6) came under scientists’ microscopes. Teachers, furthermore, were under unprecedented pressure from administrators to increase the efficiency and accountability of schools – a request characterized, in the Progressive Era, by a double emphasis on standardization and individualization.[[7]](#footnote-7) Education must have clear and objective standards, but it must also adapt to the potential and needs of each child. Fulfilling both criteria involved selecting students for the purpose of differentiated education. Until the 1910s, that selection had been operated semi-informally, using criteria such as teachers’ marks and judgment. Within the space of ten years, however, teachers’ judgment was threatened by measurement as a preferred means of assessment and selection.[[8]](#footnote-8)

To an extent, that replacement happened through a combination of favourable conjuncture and sheer force. Hypothesising about the reasons for the wide and swift spread of I.Q. tests in America, Chapman and Danziger agree that psychology professors and school administrators pressured teachers to adopt the tests. Chapman delineates the movement as ‘top-down’, with the university psychology professors at the peak of the pyramid and education bureaucrats, administrators, and headteachers serving as mid-level intermediaries. Teachers, says Danziger, lay ‘on the bottom rungs of an ever more intrusive bureaucratic structure’.[[9]](#footnote-9) Indeed the need for I.Q. testing was formulated, at first, in clear contrast to teachers’ judgments, and advertised as a replacement for them. By constantly comparing I.Q. results with teachers’ judgments, psychologists sought to expose teachers’ intuitions as less efficient and accurate.

However, the commands of professors and bureaucrats alone do not quite cover the ways in which teachers were mobilized by intelligence-testing proponents. Early on, as we shall see, that endeavour was phrased in terms of personal responsibility, not blind obedience; of spiritual and moral duty, not administrative order. At the root of it was an original flaw, which lay in teachers.

## Unreliable teachers

That flaw was the fundamental ‘unreliability’ of teachers, a keyword of a hefty corpus of texts on teachers’ marking practices, contemporaneous withthe research on intelligence-testing. The pioneer of this research was Daniel Starch, who mapped, through several studies, variations in teachers’ marks within and across different disciplines.[[10]](#footnote-10) Starch’s conclusions were sometimes striking: mathematics and geometry were no more likely than English to be marked reliably or consistently. Further work showed that teachers were poor at evaluating two pieces of work submitted by the same student, or evaluated identical work differently on different days, noted Howard Banker (who compensated for the unreferenced claim by adding, ‘These facts have been demonstrated again and again by the work of many investigators’.) [[11]](#footnote-11) The first decades of the century saw ‘much agitation’[[12]](#footnote-12) around teachers’ marks from both statisticians and psychologists: in 1918, Harold Rugg could already cite 23 articles presenting evidence of teachers’ unreliability.[[13]](#footnote-13) The accumulation of scientific evidence on the unreliability of teachers was talked-of with definitiveness and anguish:

Out of the discussion of the last fifteen years we find one point of absolute agreement, namely … (1) the striking *variability* in teachers’ marks; (2) the *unreliability*, the lack of consistency, with which teachers mark; (3) the *inconsistency in the way in which teachers distribute their marks*.[[14]](#footnote-14)To remedy these issues, suggestions ranged from publishing ‘the distribution of teachers’ marks each semester in open bulletins’,[[15]](#footnote-15) ‘training teachers to grade’,[[16]](#footnote-16) heightening self-criticality through teacher-rating scales,[[17]](#footnote-17) and, of course, standardizing marking practices and criteria.[[18]](#footnote-18) In this context, it was easy for proponents of I.Q. tests to present teachers as essentially unable to assess children’s intrinsic abilities. Since teachers could not complete the most basic task of teaching – grading consistently according to their own criteria – it was only to be expected that they would be even worse at evaluating children’s intelligence. And indeed they were; or, at least, they did less well than the Stanford-Binet scale at evaluating a child’s intelligence when checked against the Stanford-Binet scale. The claim that teachers were useless at assessing intelligence spread like wildfire in psychological writings, helped in part by the then footnote-light nature of scientific publications; references were infrequently stated, and articles, books, handbooks and manuals gave, from the early days of I.Q. testing, an impression of absolute consensus on the matter[[19]](#footnote-19). It is important to note that there was, in fact, no consensus on the validity of I.Q. testing, which was already considered at the time a polemical topic among scientists, and, as later historical work as shown, was based on highly problematic statistical calibrations, most notably to obtain a notorious ‘bell curve’ distribution[[20]](#footnote-20).

The constant comparison between teachers’ judgments and measurement made it sound like they were direct adversaries; in the battle between I.Q. test andteacher’s opinion, the latter inevitably came second. This argument deployed a simple dichotomy, still familiar today, between objectivity and subjectivity; between science and intuition. This is Brown’s argument, and she has powerfully shown how early psychologists ‘branded’ I.Q. by using ‘the cultural authority of medicine and engineering’,[[21]](#footnote-21) leaning on turn-of-the-century enthusiasm for measurement. This numbers-obsessed framework, she says, excluded educators, whose professional identity psychologists both directly and indirectly defined in opposition to their own. In Brown’s words:

More often than not, psychologists and educators were in competition with one another for power and influence in educational policy and classroom procedure… As aspiring professionals, educators stood in pathetic contrast to physicians and engineers, their accomplishments, methods, and institutions ridiculed and attacked wherever medicine and engineering were celebrated.[[22]](#footnote-22)

This rhetoric allowed psychologists to brandish scientific standards – reliability, efficiency, accuracy, and predictiveness – against which to judge teachers. Leta Hollingworth, in her handbook for teachers on gifted children,[[23]](#footnote-23) took some time to answer for her readers the question, ‘How do we know that the teacher, and not the test, is in error?’[[24]](#footnote-24) Teachers, she said, mistook beauty, poise or bookishness for intelligence; they got confused by age; and, more importantly, teachers’ evaluations did not have predictive value: ‘The research of the past twenty years has proved that the test is the best *instrument of prediction*. Children, given appropriate opportunity, usually achieve and develop in accordance with the predictions of mental tests.’[[25]](#footnote-25)

This framework defined educational method in terms always-already incompatible with the duties and the aptitudes of teachers. In 1922, arguably the zenith year of the intelligence-testing movement,[[26]](#footnote-26) with memories of another war fresh in every reader’s mind, the battle in the minds of psychologists had been so clearly won that the winning was mentioned twice by psychologist Raymond Franzen on the first page of his monograph on school adjustment for individual differences:

Standardized measurement of educational product has won its way to a recognized place in the school life of this country… The case of scientific measurement in education has been argued and won. The objections to older forms of measurement have become the criteria of the value of the new.[[27]](#footnote-27)

The ‘pathetic contrast’ between psychologist and teacher aligned with narratives of scientific progress.[[28]](#footnote-28) Scientists had at their disposal modern means of testing, ‘measuring instruments of greater accuracy than the older systems of teachers’ marks’;[[29]](#footnote-29) intelligence-testing in particular was a veritable fountain of youth: ‘The science of psychology has been vitalized and rejuvenated by this achievement’, said Herbert Woodrow.[[30]](#footnote-30) This scientific method was young, efficient, and fast: an ‘hour of examination’ was enough for the psychometricist to assess accurately what would have been inaccurately assessed by a teacher in months of teaching.[[31]](#footnote-31) Those texts rehearsed the literary trope of wonder in the face of scientific progress: ‘The teacher,’ Terman noted, ‘will doubtless be surprised that any single test requiring only ten minutes could possess this degree of accuracy’.[[32]](#footnote-32)

The ‘pathetic contrast’ was evidently gendered, too. While administrators and headteachers were mostly men, teachers of the time were overwhelmingly women,[[33]](#footnote-33) and the sexism of educational texts of the period is always apparent.[[34]](#footnote-34) Often, ‘the teacher’ is a ‘she’ – visibly not for political correctness. The teacher, ‘especially when she does not understand the situation’,[[35]](#footnote-35) could hardly be trusted with managing all levels of ability in a classroom. Terman himself, despite purposefully recruiting an all-female group of researchers,[[36]](#footnote-36) felt that he had to warn teachers that they should not babble about children’s IQ results.[[37]](#footnote-37) The intuitive, gossipy female teacher is a constant presence in the writings of the time. Thus the ‘pathetic contrast’ between teachers and scientists, and the ‘ridiculing’ and ‘attack’ of their methods, did exist; spoken about under the guise of scientific neutrality, teachers were belittled and deemed unfit for the task.

However, Brown’s ‘pathetic contrast’ suffers important nuances. As I now go on to discuss, in both the discourses and practices of intelligence-testing proponents, the explicit denigration of teachers constantly coexisted with placating declarations of unity, and offers of help and redemption.

## All too human: Diplomatic addenda

The modern concept of intelligence belonged – because they had invented it – to psychologists. This meant that the literature on intelligence could be less territorial than the literature on marking. When Terman’s handbook *The Measurement of Intelligence* came out in 1916, he could easily state that the teacher ‘has no means of knowing whether the average for her class is above, equal to, or below that for children in general. Her standard may be too high, too low, vague, mechanical, or fragmentary’; or that teachers ‘don’t have a very definite idea of what constitutes intelligence’.[[38]](#footnote-38) This was not ruthless denigration of teachers, but logic: ‘what constituted intelligence’ could only be calculated using tests; intelligence had never truly belonged to teachers’ expertise anyway. Psychologists were not truly chiding teachers for their ‘grave errors’ in assessing intelligence, nor suggesting that this made them bad teachers. The obvious foreignness of the concept of intelligence to teaching placed psychologists from the start in positions of pedagogues.

Teachers’ ‘failures’ in that domain were thus often mitigated by pacifying claims: particularly, the common assertion that teachers were no more useless than anyone at assessing intelligence. Hollingworth highlighted the difficulty, ‘even for those who have devoted years of special study to this subject, to arrive at a perfectly clear concept of intelligence’[[39]](#footnote-39). Terman again: ‘Every psychologist who is experienced in the mental examination of school children knows that his own or the teacher’s estimate of a child’s intelligence is subject to grave and frequent error’.[[40]](#footnote-40) The *humanity* of teachers – and psychologists – was their downfall, their ultimate limit: (un?)fortunately, ‘teachers [were] not automatons’. [[41]](#footnote-41) Particularly noted were the errancies of everyday life: teachers’ judgements could be ‘dulled’ by ‘mental distraction’, poor sleep or personal issues.[[42]](#footnote-42) ‘They are the kind of errors which any person, however discriminating and observant, is likely to make in estimating the intelligence of a subject without the use of standardized tests,’ concluded Terman.[[43]](#footnote-43) The matter was out of the hands of common mortals.

To highlight this reinforced the notion – crucial, I think, to the spread of intelligence-testing within schools – that I.Q. tests were essentially *different* from, rather than a superior species of, teachers’ judgment. This distinction allowed psychologists to precisely locate measurement *in relation to* – rather than as replacement for – the teacher’s work. And that location was unequivocally ‘*before*’: the intelligence test preceded, supported, grounded the teacher’s work. In Raymond Franzen’s words: ‘Measurement should form the basis for teachers’ opinion; it should neither supplant nor supplement it’. [[44]](#footnote-44)

Otherwise put: it should come *first*. Psychologists insisted on strong sequential order in the uses of science and teaching. Hollingworth, in the 1930s, would stress with particular force the necessary anteriority of scientific testing over educational processes:

Before education can discharge this most important task of all with economy and justice, it must become a science. The science which is fundamental to education is psychology. Psychology had to develop the methods of mental measurement before there could be accurate or humane dealing in a system of compulsory education. We must take “the measure of a man” before we can know how to educate him.[[45]](#footnote-45)

*Before, before, before*: intelligence-testing, arrived recently on the educational scene, was positioned retroactively by its advocates as a foundation. I.Q. tests were the entry-point into the child’s life, providing the base setting for education.

This claim helped negotiate spaces, or moments, for *both* teacher and psychologist in the educational system, allowing psychologists to avoid alienating a much-needed potential groundforce. However, it also established a one-way system of favours between the two professions, with psychologists as benefactors and teachers as benefiting. Knowledge flowed in just one direction. Measurement could inform teaching, but the reverse was not true: even when they held anecdotal validity, teachers’ judgments were not useful to psychologists. Teachers studied by Binet said that they could ‘tell’ a child’s intelligence by ‘a glance of the eye’ – but, asked Terman rhetorically, ‘how are we going to *standardize* a “glance of the eye or an “expression of curiosity” so that it will serve as an exact measure of intelligence?’.[[46]](#footnote-46) The positioning of I.Q. as anterior to teaching – not an eviction, but a *grounding* of teaching – was a powerful way of establishing it as a help or as a gift, not as a replacement. Teachers’ all-too-human unreliability, it was understood, could be redeemed by adopting the tests. And in that respect they would get a lot of help.

**Helping teachers to help themselves**

The purpose of this booklet is to make the Binet-Simon scale available in popular form to the great army of men and women active in public and semi-public positions in the care and control of childhood.[[47]](#footnote-47)

‘Teachers must learn to use tests,’ said Terman, in bold.[[48]](#footnote-48) That was an accepted fact, though it is unclear when that fact actually became accepted; it just was, thanks to lavish use of what one might call arguments from obviousness. Texts of the time overused adverbs such as ‘clearly’, ‘obviously’, ‘evidently’ when asserting the power of science.[[49]](#footnote-49) The obvious remedial legitimacy of intelligence measurement became a leitmotiv, relying also on statements of exclusivity: ‘The only way of escape from the fallacies we have mentioned lies in the use of some kind of refined psychological procedure’;[[50]](#footnote-50) ‘Without the use of scales for measuring intelligence we can give no better answer as to the essential difference between a genius and a fool than is to be found in legend and fiction’.[[51]](#footnote-51) Handbooks for teachers on mastering I.Q. testing routinely placed such claims at the start, stating, rather than providing evidence for, the normalisation of tests in schools. Thus Terman, in his introduction to *The Intelligence of Schoolchildren*, asserts, ‘That [the tests’] universal use in the schoolroom is necessary to educational efficiency will doubtless soon be accepted as a matter of course’.[[52]](#footnote-52) Perhaps he need not have bothered, because his eminent contemporary and Dean of Stanford School of Education, Ellwood P. Cubberley, had already prefaced the book by saying, ‘The large usefulness of intelligence testing has been thoroughly demonstrated… The question now is not, Are intelligence tests of value? But, How may teachers and principals be made masters of their use?’[[53]](#footnote-53)

How indeed? The following, by Buckingham, can serve as a guide:

The psychologist and the teacher are getting together on this question. The psychologist has simplified his methods of testing and he is training the teacher to use the simpler instruments. On the other hand, school people are realizing with conviction that they need to know and to know in terms of the new methods of measurement, the physical and mental development of the children entrusted to their care.[[54]](#footnote-54)

This quotation rehearses three major narratives in psychologists’ texts of the time: first, the creation of a sense of camaraderie between psychologist and teacher; then, the empowerment of teachers through the use of simplified scientific instruments; and finally, the assertion of teachers’ gratefulness. I analyse them in turn below.

## All in this together

First, Buckingham’s quotation gives a sense of comradeship, of ‘togetherness’, between teachers and psychologists in the endeavour of school improvement. This ‘togetherness’ had many declensions. As above, it was partly an assertion of shared humanity, humble in the face of awesome science, and reclaimed a common social and philosophical project; teachers were *de facto* assumed to share psychologists’ goals regarding student selection.

That claim was not solely philosophico-political; it also leaned on a recent history of increased collaboration between teachers and psychologists. In previous decades, the work of Stanley Hall and of the Child Study movement had supported a practice that would prove essential to the spread of intelligence-testing: teacher-led research.[[55]](#footnote-55) Teachers of that generation, well-formed, trained in psychology, were, psychologists claimed, ‘themselves becoming active, creative workers in the field… [carrying] on scientific studies not only in the practice schools… but also in the school systens [sic]’,[[56]](#footnote-56) and reporting to education journals. In the ideal worlds of Hall’s followers – and, certainly, in Terman’s mind – teachers had huge potential as testers, data gatherers and day-to-day supporters of their cause in schools.

Evidently, psychologists’ vested interest in the spreading of tests and method must have created a strong incentive to self-publicise; Terman especially had clear economic stakes in the matter.[[57]](#footnote-57) As Zenderland has shown, Terman worked hard to stifle his colleagues’ criticisms of teacher-testers: it was essential for him to present and train teachers as legitimate testers. In a striking early article, Terman lambasted the ‘bombast, pretentiousness and shockingly open claims to recognition’[[58]](#footnote-58) of those who, such as J.E. Wallace Wallin, had cast doubts on the appropriateness of encouraging teachers to use intelligence tests.[[59]](#footnote-59) Psychology must show a united front, and profess and practice a collaborative enthusiasm that left little space for contestation or conflict.

*‘The teachers protested’*

However, although teachers, administrators and psychologists were indeed ‘getting together’ increasingly often, that repeated claim uneasily covered up a significant problem with I.Q. testing: the diversity between teachers in their acceptance of the practice. Some teachers enthusiastically adopted tests, but the warmth was not universal. In the following, Terman recounts his difficulties with getting kindergarten teachers to evaluate children’s intelligence:

When asked to estimate the intelligence of each child on the usual scale of five: very superior, superior, average, inferior, and very inferior, the teachers protested that there was almost nothing in kindergarten work on which they could base a judgment. The ratings on intelligence were finally secured, but they correlated with IQ’s only to the extent of .29.[[60]](#footnote-60)

Kindergarten teachers, in fact, would have been particularly likely to be reticent to I.Q. tests, being, as they were, the most child-centred corps of the teaching body and quite opposed to the rationalization of education.[[61]](#footnote-61) But Terman’s tale of the protesting teachers was a more general warning to those who sowed dissent, and it was also an implicit call to administrators to support the work of psychologists, addressing, ‘over the heads’, so to speak, of its primary of audience of teachers, the corps of principals and superintendents who guaranteed unity on the ground. Psychologists were greatly aware of the importance of that cementing layer: the school administrator was everywhere mobilized to motivate troops of teachers. But very rarely was there any reference to forceful imposition of intelligence-testing; instead, superintendents and principals were asked to be motivational speakers, explainers, and distillers of I.Q.’s associated socio- educational vision.[[62]](#footnote-62) In this extract from an article by J. Carleton Bell, we see the extent to which teachers’ support, as envisaged by psychologists, depends on deeply emotional engagement with the cause, fuelled by a kind of magical thinking set against narratives of scientific progress:[[63]](#footnote-63)

The recognition of the value of measurements for the teacher can only come about through the encouragement of the supervising staff. When superintendent, principal and teacher unite whole-heartedly to study the needs and capacities of individual children and to base the school work upon their findings, much of the drudgery and grind of teaching will disappear, and educational measurements will attain their greatest usefulness.[[64]](#footnote-64)

But not all teachers and, indeed, superintendents aligned with these wishes. Unlike university psychologists and administrators, teachers had to deal with the everyday consequences of I.Q. testing, including parental reticence to see their children tested.[[65]](#footnote-65) While the evidence is sparse, there is reason to believe that teachers did not unanimously comply with the pressure to test. Judith Raftery, in a persuasive case study of teachers’ uses of I.Q. tests in Los Angeles in the 1920s and early 1930s, shows the confusion and uncertainty generated by the tests.[[66]](#footnote-66) She also highlights that teachers were themselves adapting, tweaking or translating tests, or used achievement rather than intelligence tests, which they saw as more useful to their work. Wayne Urban’s study of the National Education Association maps the resistance of teachers to the desire for control coming from upper echelons, and he notes that superintendents’ compliance was not always matched with teachers’ enthusiasm.[[67]](#footnote-67) Lana Goldstein’s study of the Chicago Teachers Federation, and Ellen Condliffe Lagermann’s work on the Progressive Education Association, similarly show the considerable defiance of unionised teachers towards I.Q. testing for the purposes of tracking.[[68]](#footnote-68) Teachers among themselves and in professional conferences regularly denounced intelligence tests, and clearly some superintendents did too. See for instance, in 1931, *The New York Times’* report on the New York State Teachers’ Association’s annual meeting:

Addressing 500 elementary principals of the State at a section conference, Dr Edwin C. Broome, Superintendent of Schools in Philadelphia and president of the National Council of School Superintendents of the United States…assailed intelligence tests as an inadequate measure of a child’s ability. … Listing as one of the ‘unsolved problems’ the search for reliable measures of children’s intelligence, Dr Broome said intelligence tests ‘are not the final word in the segregation or classification of boys and girls’. [[69]](#footnote-69)

Thumbing through *The New York Times* at the time, the lay reader would have encountered some quite intense articles denuncing psychologists as ‘[doing] little but give intelligence tests’[[70]](#footnote-70), or pointing at ‘an army of young Ph.D.’s who have never taught children framing up these tests and judging the efforts of teachers by them’.[[71]](#footnote-71) Psychologists’ defensiveness and assertions of ‘togetherness’ indicate that they knew that teachers were a too imperfectly united ‘army’. In 1923 in *The* *New York Times*, in an article rather clumsily entitled ‘No Fundamental Basis for Disagreement Among Teachers’, Virgil E. Dickson vehemently smothered the controversy, asserting of the teacher, ‘She wants something that will help her to understand better the nature and ability or present condition of each child and something that will enable her to know better how to teach him. She has a feeling that these tests will help her.’[[72]](#footnote-72) The much-decried intuitiveness of teachers was here, ironically, reclaimed as a testament to the tests’ usefulness.

Yet there is no contesting the use of intelligence tests in schools overall. Goldstein reports that in 1925, 64% of elementary schools were using I.Q. tests for ability grouping.[[73]](#footnote-73) And even when there was resistance to the practical aspects of testing, it does not follow that there was resistance to the principles behind testing. Raftery wisely warns against the temptation to take ‘eminently quotable leaders’ such as Terman at their word regarding ‘how school testing worked’[[74]](#footnote-74); but her study does show teachers who sought to modify or adapt tests in part because they hoped for a more perfect test. As such, they aligned with psychologists in their dreams, doubtlessly very sincere, of genuine improvement to the lives of thousands of children.[[75]](#footnote-75)

## ‘For the benefit of the teacher’

In this endeavour, the psychologist was to define the tests, as Dickson’s previously-quoted article stated, ‘for the benefit of the teacher’. The relation between the two professions was never presented as equal: teachers had something to gain from it, while scientists had only duties. Teachers, therefore, were everywhere *used*, but nowhere treated, addressed, or talked about, as *useful*. Psychologists’ both rhetorical and practical mobilizations of teachers skirted around acknowledging the actual services rendered by their soldiers, foregrounding instead psychology’s favours to education.

Characteristically, the quotation by Buckingham presents science as having generously empowered teachers by making its instruments accessible to them. And indeed, to a great degree, it had. Handbooks and training courses materialised psychologists’ intentions that teachers should learn to use tests properly, and articles or books described in detail the necessary courses: training, Terman suggested, took half a year; trainees had three lessons per week and additional courses in educational psychology; teachers already in exercise could take Saturday or afternoon classes and summer courses.[[76]](#footnote-76) Teachers’ colleges and normal schools were a priority: in accordance with the narrative of chronological precedence of I.Q. tests over teaching, it was particularly important to cultivate in young people a taste for the practice.

Though teachers were indeed furthering psychologists’ enterprise, doing research, spreading information about the tests, administering them, endorsing their values, and getting schools to buy the materials,[[77]](#footnote-77) psychologists’ texts avoided acknowledging that teachers were rendering any kind of service; instead, they drew much attention to the fact that psychologists had served teachers in the first place by simplifying tests.[[78]](#footnote-78) Simplicity was a unique selling point of I.Q. tests[[79]](#footnote-79) and also doubtlessly facilitated their adoption; the concept was simple, the test was simple (the scoring was not so simple; more on that later). The insistence on simplicity implied two things, which I detail below: that teachers were not able to grasp very complex ideas, and that simplifying things for teachers was the psychologists’ duty.

First, there existed, from the early days of intelligence-testing, much anxious literature on teachers’ intelligence; when ‘bineted’, normal school students and teachers in exercise were found to have disappointingly average or even below-average I.Q.[[80]](#footnote-80) The ‘facts’ were superficially deplored, and presented as debunking the myth of the all-knowing teacher.[[81]](#footnote-81) But, having dismissed the infallible teacher figure, psychologists recast it into a much stronger ally, the *perfectible* teacher figure. Teachers were now addressed, talked about and treated as lifelong students. Psychologists relished presenting teachers as pupils, or even as children: ‘Before school marks can have a definite meaning there must be a standardized measure developed; and teachers must be taught to use it, just as children are taught to use the foot rule’, says Sharp.[[82]](#footnote-82)

This was not just rhetoric: teachers were effectively finding themselves increasingly taught, tested and controlled by higher authorities, both administrative and scientific.[[83]](#footnote-83) Handbooks and manuals for teachers employed precisely the pedagogical strategies that were being extolled for use with students: Edward Strong’s 1920 *Introductory Psychology for Teachers*, for instance, highlighted in its introduction the importance *for teachers* of learning by doing and of ‘starting with concrete experiences of everyday life’, and the handbook used storytelling to introduce psychological knowledge.[[84]](#footnote-84) The double emphasis on teachers’ ‘low’ intelligence and on further education created teachers who were permanently in training, waiting for the next scientific discovery.

Relatedly, psychologists presented their own work as going beyond the call of duty. In their view, teaching teachers and adapting their scales had turned psychology into an especially charitable, superegatory science. The ‘services’ rendered by psychology were everywhere emphasised, from ‘large service to the teaching staff’ to ‘[serving] the children of Berkeley’[[85]](#footnote-85) through to ‘services to the human race’[[86]](#footnote-86) and equally abstract assertions of ‘contribution’ to ‘the science and the art of education’.[[87]](#footnote-87) Still, the practical sides of those services were never eclipsed: tests were ‘the teacher’s useful tools’, which allowed the teacher ‘in the long run less work, and much easier and more successful work’[[88]](#footnote-88). I.Q.-testing, both as philosophical enterprise and as practical tool in the day-to-day running of schools, was for psychologists a calling, for teachers a blessing and a wealth of opportunities.

*Grateful teachers*

So they must be grateful for it. In the earlier quotation by Buckingham, the psychologist ventriloquizes teachers, asserting both their indebtedness and their support: ‘school people are realizing with conviction that they need to know and to know in terms of the new methods of measurement, the physical and mental development of the children entrusted to their care’[[89]](#footnote-89). That was common practice, especially in handbooks destined to teachers, where fictional teachers’ voices were routinely heard, and fictional popoulations of teachers asserted or admitted the usefulness of tests.[[90]](#footnote-90) Real teachers were also quoted in intervention studies in schools, beginning with Dickson’s pioneering work.[[91]](#footnote-91) Dickson, a mentee of Terman,[[92]](#footnote-92) is credited with achieving the earliest large-scale implementation of I.Q.-testing in American schools, in the public schools of Oakland, California, in 1917. Dickson’s *Report* on the experiment, published in 1918, begins with a subpart on ‘The Interest in Research Work Aroused Among Teachers and Principals’; Dickson’s first concern had been to survey teachers and administrators to map their needs. Dickson concluded that teachers, although very willing to accept school readjustment by selecting students through I.Q.-testing, were ‘very helpless unless they have the co-operation of the administrative forces. Hence the class for principals was organized that we might, all working together, study these problems.’[[93]](#footnote-93)

We see here an early example of the rhetoric of togetherness developed above, as well as a clear insistence of the contribution of psychology; but Dickson, not content with his sole surveys, added extra ‘evidence’ on the helplessness of teachers: ‘Great interest has been shown by both teachers and principals with the result that the Research Department has been flooded with calls for help which it could not find time to give.’[[94]](#footnote-94) This is, I have found, a typical move from psychologists in that era; they constantly alternated between quoting real teachers and ventriloquizing fictional teachers, without it being always clear how many teachers had said what to whom. In a later work, Dickson affirmed that ‘Teachers who have taken the work have expressed their appreciation of the instruction received and the benefits that have accrued from it’[[95]](#footnote-95), and in the same text, ‘Every teacher who is trained in the principles and practice of mental testing finds it an invaluable aid in her classroom teaching’[[96]](#footnote-96), mingling first-hand reports with implausible generalisations. This artistic blurriness, at odds with the psychologists’ insistence on fact and data as basis for intervention, occurred in many texts of the period.

# Controlling the teacher-tester

*No liberties*

The rhetorical and practical set-up of I.Q. testing as a one-way favour from psychologist to teacher provided an ideal ethical framework within which to enclose and police teachers’ uses of tests. For if teachers were to be ‘experts’ of a kind in assessing children’s intelligence, it was important that psychologists retained control over their uses of the tests. Early on, there was some criticism, internal to psychology, surrounding teachers’ access to intelligence-testing. In 1922, , Marion Trabue penned a piece warning of the ‘dangers’ of teachers’ inexpertness when administering tests.[[97]](#footnote-97) That critique is particularly intriguing for its intermingling of ‘pro-teacher’ ideas – tests, said Trabue, should ‘never be substituted for common sense and good judgment’[[98]](#footnote-98) – and calls for exclusive use of scientific methods by scientists, for fear of abuse. Regulatory anxiety thus led, paradoxically, to greater recognition of the domain of expertise of teachers: teachers gained ‘common sense’ and ‘good judgment’ when commentators estimated that it was necessary to restrict their uses of tests.

But most psychologists were absolutely adamant that teachers should indeed be testers. Only, that ‘favour’ was to be strictly controlled. The psychologist, Prometheus-like, had brought to teachers a new light; the least teachers could do was take their responsibility seriously and not overstep their role. Teachers-testers must be disciplined, focused and motivated, like a regiment: Terman particularly liked the expression ‘the rank and file of teachers’.[[99]](#footnote-99) I.Q. testing, teachers must understand, suffered no imprecisions. Terman allowed for the possibility of a teacher in practice having to teach herself how to use tests (an evident sign of psychologists’ need for teachers). That was not ideal, but better than not using tests at all. In that case, she had to ensure she ‘studied the directions with conscientious care and [learned] not to take liberties with them’.[[100]](#footnote-100)

In order to ensure that no ‘liberties’ were taken, becoming a certified Bineter involved strict procedures of control. Scoring the tests was the most difficult part of the process; teachers’ scores had to be checked by psychologists ‘until the right habits have been thoroughly established’.[[101]](#footnote-101) Guy Whipple details the procedure:

The tests should first be scored by the teachers, and that all papers should then be shipped to the office of the chief examiner and checked or rescored by clerks trained for this work. If the first half of the papers returned by a given teacher are found to be without error, the remainder may be regarded as sufficiently certain to be without systematic errors.[[102]](#footnote-102)

Dickson had more drastic measures in mind: after fifteen weekly lectures, teachers’ scoring aptitudes on sixteen to twenty tests were checked, after which ‘the teacher [was] henceforth known as a “certificated mental examiner”‘; but,

Even at this point, however, the training in the work of individual mental testing is not complete. Tests submitted by a certificated examiner are still checked from time to time and errors are called to the attention of the teacher concerned.[[103]](#footnote-103)

Teachers were not allowed to forget who had trained them, and their own fallibility; they were always at risk of error, always *in progress*. Those practices and discourses glossed over a fact which was, from the early 1920s, increasingly uncomfortable to psychologists: the unreliability of scoring on Binet and other scales.[[104]](#footnote-104) The obsession over whether teachers were following strict procedure raises questions about the validity of a test advertised as simple; why would it leave so much space for individual error? Psychologists turned that awkward question into a way of justifying control over teachers-users. Thus they maintained ownership of methods, knowledge, and, of course, qualifications.

For no liberties, either, were to be taken by teachers regarding their true rank. The teacher, Terman sternly notes, ‘must understand clearly that the mere ability to give a Binet test acceptably gives her no claim to the title “clinical psychologist”.’[[105]](#footnote-105) In an early text, while defending the legitimacy of teachers as testers, he warns against ‘confusing the complex and baffling science of clinical psychology with the mechanical use of the Binet intelligence tests’.[[106]](#footnote-106) Elsewhere he recommended that teachers avoid offensive vocabulary such as ‘dullard’ of ‘feeble-minded’ – words that he himself lavishly used.[[107]](#footnote-107) Clearly, not everything that psychologists allowed themselves was allowed to teachers. And even though teachers gradually began to publish in educational psychology journals, reporting, for instance, on the experimental classes set up by psychologists,[[108]](#footnote-108) there was always an editor somewhere to warn that teachers were not the most accurate of researchers.[[109]](#footnote-109)

Within this system of favours, the teacher, insider-outsider, used-user, agent-puppet, ally-adversary, helped define – but without ever being allowed to *defy* – psychology as profession and science. Teachers constituted for psychologists a kind of outer limit – an indirectly definitional *alter ego*, necessary to their cause but kept at arm’s length from their knowledge. Psychologists, throughout the era, glazed their practices and knowledge with the impermeable veneer of charitable pedagogy.

## Honour and shame

Easily detectable in psychologists’ tight control of teacher-testers is the intrinsic distrust of the teaching profession by scientists of the Progressive Era, who, perhaps not without reason, tended to view teachers as ‘great conservatives’.[[110]](#footnote-110) The political leanings and activism of teachers in the Progressive Era have been debated,[[111]](#footnote-111) but there seemed to have been little doubt, in the minds of intelligence-testing proponents, that teachers did not share their progressive enthusiasm.

When they were cast as either allies or adversaries of the cause, it was often along political lines. Teachers who embraced the tests were attributed the favourable label of progressive: ‘Progressive educators are no longer interested in arguments regarding the validity of the test method’, said Terman;[[112]](#footnote-112) ‘The interest shown among principals and teachers … is indicative of open minds and a progressive attitude’, added Dickson.[[113]](#footnote-113) The definition of a good teacher incorporated indebtedness to scientists: ‘The efficient teacher is the one who is accepting all the help which tests can give in promoting, drilling, and guiding’.[[114]](#footnote-114) Promotion for those good soldiers took the form of certificates (not to be confused, as pointed above, with a psychology degree), but mostly it involved spectacular professional improvement: the ideal teacher, touched by the wonders of measurement, was ‘transformed from a tiresome drillmaster into a guide to the child’s best development’;[[115]](#footnote-115) similar metaphors of metamorphosis surrounding teachers abound in literature of the time.

But those miracles did little to hide the fact that the ‘progressive’, ‘efficient’ ideal teacher-tester was mostly an impeccable executive. That teacher had abandoned the pipe dream of improving children through education, and embraced the ‘fact’ of limited potential; she had reoriented her efforts towards selection. In a striking editorial to the *Journal of Educational Research* in 1921, Buckingham vociferously asserted the uselessness of teaching.[[116]](#footnote-116) Schooling, he argued, is solely valuable as a selective enterprise: ‘We cannot create capacity’,[[117]](#footnote-117) only identify the maximum potential of students and teach them to the brim. Within this drastically restricted understanding of education, intelligence-tests, said Buckingham, allowed teachers to select more accurately than ever. Only then could the transformation occur: ‘Thus, the teacher becomes a guide and director,’ Buckingham insisted; ‘He is no longer blind’.[[118]](#footnote-118) He concluded, ‘When to the old consecration by which the craft has been distinguished is added the new insight which modern methods can confer, teaching will acquire new dignity’.[[119]](#footnote-119) That final flourish implied that psychology had, yet again, bestowed upon educators nobility and validation. It also gave teacher-testers little visibility beyond the horizon of the Binet test.

Indeed, psychologists talked little about what teaching actually involved except testing and sorting students into different tracks according to their I.Q. results. This put teachers in the odd situation of having knowledge and techniques that ennobled them, but did not lead to very precise applications. Yet this was the teachers’ problem. Buckingham did not believe that teachers could do much more than select students; but paradoxically, in an earlier editorial, he had railed against teachers who complained that the obsession for measurement had not yet improved teaching:

If instruction has not been improved as the result of knowledge gained from testing, it is not the fault of the tests, for they have provided the facts and to provide the facts is their sole office. … It is primarily the fault of the school people. … It is the school people who must be heard from. The “next step” after testing is theirs.[[120]](#footnote-120)

The problem for psychologists in envisaging a ‘what next?’ to teachers’ uses of tests may simply have been caused by their realisation that, indeed, that territory belonged to educators. It may also have come from fifteen years of meticulous focus on measurement rather than application. But, I would argue, the vagueness of the educational horizon beyond testing and sorting mostly illustrates the difficulty for the discourses of intelligence-testers to solidify into specific, clearly spelled-out programmes for the future that teachers could take up. Teachers hovered, as we have seen, in a semi-restricted space that precluded ownership of the knowledge. The ‘good’ teacher – the certified Bineter, the believer, the good soldier – was inevitably also a bad teacher; she could not convert into practice a knowledge that had never truly been her own. Teachers, having been enlisted, addressed, trained and talked about mostly as executives, were the ideal people onto whom to diplace complaints about lack of effective school improvement. Their lack of success also hinted that the originary debt had not been repayed – but arguably, within those parameters, it could not ever be.

Meanwhile, teachers who gave signs of abusing, misusing, or simply not using the tests exposed themselves to public shaming, in the form, for instance, of Hollingworth’s legendary use of snarky scarequotes: ‘The idea that [gifted] children exist at all is even laughed to scorn by teachers and principals who have a quarter of a century of “experience” behind them. … They become known only to those educators who “believe in” mental tests.’[[121]](#footnote-121) Slightly more menacingly, they were treated as unpatriotic, undemocratic, and unwilling to participate in the progress of the nation. Thus Dickson:

We are fighting to make the world safe for democracy. We must fight in a different way to make education and vocation recognize individual differences that will make our own people safe for democracy.[[122]](#footnote-122)

And Alexander Carter, in an article on ‘Presenting educational measurements so as to influence the public favourably’:

The measurement energy so far has mainly attacked problems of inner development. The members of this association of course desire to be true educational patriots. To attain this wish, they must do their fair share of the fighting necessary to secure better school support.[[123]](#footnote-123)

Despite their martial rhetorics, those two quotations are less aggressive than the ample corpus of texts which denounced the lack of collective ethos and quasi-criminal streaks, of teachers who failed to adhere to the cause of progressive education. This was especially the case in relation to giftedness;[[124]](#footnote-124) those teachers who neglected to use appropriate tools to detect children of high intelligence were no less than modern Frankensteins: ‘It is possible through wrong education to turn a gifted child into an intellectual monstrosity’, said Howard Taylor.[[125]](#footnote-125) The teacher impervious to, or unwilling to accept the help of, the psychologist, risked creating lazy, antisocial children, whose energies were redirected towards ‘Bolshevism’.[[126]](#footnote-126) In short, the acceptance of I.Q. tests was in itself a test of true democratic spirit; socialist radicals, or unrepentant reactionaries, lurked inside those who refused to use the tests.[[127]](#footnote-127)

# Conclusion: ‘The only distributing agency’

I will conclude by noting the invisibility, in scientific writings of the period, of the fact that teachers, too, were doing an immense service to scientists and to psychology. Psychologists hardly ever thanked teachers; when they praised them, it was in deliriously grand terms, equating the ideal teacher with ‘Jesus’ or ‘Plato’.[[128]](#footnote-128) But scientists rarely expressed gratitude for the day-to-day efforts of ordinary teachers or for their contributions; they focused, instead, on teachers’ gratefulness towards them. It is not that teachers’ contributions, initiatives and transgressions did not exist, or even that they necessarily went unnoticed among individual psychologists. It is, I contend, that the discursive framework simply did not allow for thanks to be given to teachers, because teachers’ and psychologists’ works were from the start precisely not conceptualised as an *exchange* of favours.

Here it is useful to recall that elsewhere in educational research, contrasting frameworks existed. Lagemann’s history of educational research opens with an originary battle between John Dewey and Edward Lee Thorndike – won, she argues, by the latter – that led to the perception of schooling as an ever-improvable system.[[129]](#footnote-129) Dewey had the words to name teachers’ contributions, because he had a theoretical framework within which they made sense: he portrayed education ‘as an art’, which does not ‘turn [its back] upon science’, yet ‘which happens not because of scientific method but because of departure from it’.[[130]](#footnote-130) Working from the premise of the irreducible unpredictability of teaching situations, Dewey highlighted the importance of the ‘*other* things’[[131]](#footnote-131), the interstitial spaces of knowledge and questioning – mismatches between theory and practice, differences between measured and observed achievement, problems that ‘arise in actual relationships with students’[[132]](#footnote-132). The intelligent detection and investigation of those glitches was not just the teacher’s domain but the point of origin of scientific enquiry. Within that framework, Dewey could easily conceptualise teachers’ initiatives, actions *in situ*, and unique contributions to educational science. Not so, as we have seen, within the discursive framework of intelligence-testing proponents.

Educational journals like the *Peabody* sometimes reassured their readers as to the value of their work: in 1928, Bruce Ryburn Payne, mapping ‘Ten years of progress in the service rendered by normal schools and teachers colleges’, dropped a casual reminder:

We educators have ourselves had so much talk about buildings, equipment, laboratories, libraries, statistical work, and this, that, and the other, in recent years, that somehow even we seem to have forgotten that the teacher is the only distributing agency.[[133]](#footnote-133)

Some of the reasons behind that ‘somehow’, I hope, have been unfolded in this article. The invisibility of the teacher as ‘distributing agency’ was certainly in part the fruit of an era where increased standardization and bureaucracy in education had somewhat atomized teachers’ roles and responsibilities. But that was not the only factor. In fact, discourses surrounding teachers in respect to intelligence-testing overemphasised the roles, responsibilities and motivation of teachers; the I.Q.-testing movement was extremely teacher-focused. Yet the celebration of scientific rigour, of inter-professional camaraderie and of the favours rendered by psychology to education meant that teachers’ contributions found few referential points by which to be defined in other ways than as due repayment for services rendered. Thus, whether actual or potential, teachers’ contributions were not legible themselves as a help on the way to educational ‘progress’. Beyond its specific aspects, this particular story in the Progressive Era can help us think more generally of the ways in which teachers’ services to society or science might be rendered unthinkable, or eclipsed, when framed within systems of thought that demand their existence, yet deny them gratitude. The assumption of a thankfulness going only one way was, I think, a key ‘weapon’ of I.Q. testing proponents in the ‘war’ with (and alongside) teachers in the Progressive Era. The relationship was characterised by an assertion arguably stronger than it would have been if scientists had repeated *ad infinitum* that they were more accurate, more rational, more efficient than teachers. That assertion was that psychologists deserved the thanks.

1. This was a dominant claim in 1960s and 1970s approaches to the history of intelligence testing. For an example of claims that teachers were being systematically ‘colonized’, see Richardson, Theresa, and Erwin V. Johanningmeier. “Intelligence testing: the legitimation of a meritocratic educational science.” *International Journal of Educational Research* 27, no. 8 (1998): 699-714. Later histories of teaching, especially with a feminist slant, have called for an end to the view of the Progressive Era ‘as a time when teachers were reduced to little more than pawns in large-scale school systems’ (MacDonald, Victoria-Maria. “The paradox of bureaucratization: New views on progressive era teachers and the development of a woman's profession.” *History of Education Quarterly* 39, no. 4 (1999): 427-453, 432). Yet the processes of belittlement and objectification of teachers historically are still persuasively discussed and denounced today. In this journal, one can think of Kate Rousmaniere’s sophisticated study of the rhetorics of disability in relation to teaching (“Those who can't, teach: the disabling history of American educators.” *History of Education Quarterly* 53, no. 1 (2013): 90-103), or of Jonna Perrillo’s no less enlightening Foucauldian-feminist approach of the discursive control of teachers’ bodies in the interwar era (“Beyond “progressive” reform: Bodies, discipline, and the construction of the professional teacher in interwar America.” *History of Education Quarterly* 44, no. 3 (2004): 337-363). [↑](#footnote-ref-1)
2. Chapman, Paul Davis. *Schools as sorters: Lewis M. Terman, applied psychology, and the intelligence testing movement, 1890–1930*. New York University Press, 1988. [↑](#footnote-ref-2)
3. Brown, JoAnne. *The definition of a profession: The authority of metaphor in the history of intelligence testing, 1890-1930*. Princeton University Press, 1992. I use the term ‘I.Q.’ interchangeably with ‘intelligence’. See also Danziger, Kurt. *Naming the mind: How psychology found its language*. Sage, 1997. [↑](#footnote-ref-3)
4. The use of war lexicon, throughout this article, winks to Dana Goldstein’s popular account of the history of teaching (Goldstein, Dana. *The Teacher Wars.* New York: Anchor, 2014). War metaphors, perhaps cliché in historical research, are not an abuse of language when discussing ‘America’s most embattled profession’, and martial rhetoric was amply deployed by psychologists of the Progressive Era. [↑](#footnote-ref-4)
5. See Chapman’s excellent account (*Schools as Sorters*), and Tyack’s classic study of the rise of urban education in America (Tyack, David B. *The one best system: A history of American urban education*. Harvard University Press, 1974). For contemporaneous accounts, Judd, Charles Hubbard, “The influence of scientific studies in education on teacher‐training institutions.” *Peabody Journal of Education* 2, no. 6 (1925): 291-300; Tansil, Rebecca C. “Steps in the history of standardization of normal schools and teachers colleges.” *Peabody Journal of Education* 7, no. 3 (1929): 164-167. [↑](#footnote-ref-5)
6. Chapman, *Schools as Sorters*, 35. [↑](#footnote-ref-6)
7. A dichotomy articulated by most historians, such as Chapman, *Schools as Sorters*; Danziger, *Naming the mind*; Tyack, *The one best system*. [↑](#footnote-ref-7)
8. The genesis of intelligence-testing in the United States is extremely well-known; I sketch here a summary as backdrop. In the early 1900s, French psychologist Alfred Binet, mandated by the French government to study ‘feeble-minded’ children, developed with Théodore Simon general intelligence tests based on a distinction between mental and chronological age. American psychologists Goddard and Terman enthusiastically adopted and adapted the scale; Terman’s version, the Stanford-Binet scale, was published and taught in books, articles and handbooks for teachers in the mid-1910s. Schoolchildren were ‘bineted’ on ever-larger scales throughout the 1910s and 1920s. In 1919, a team comprising Terman, Yerkes, Haggerty, Thorndike and Whipple developed the National Intelligence Tests, of which 200,000 copies were sold in 1920. See for an excellent overview Zenderland, Leila. *Measuring minds: Henry Herbert Goddard and the origins of American intelligence testing*. Cambridge University Press, 2001; for a briefer history, Fass, Paula S. “The IQ: A cultural and historical framework.” *American Journal of Education* (1980): 431-458; for a critical history, Kamin, Leon J. *The science and politics of IQ*. Psychology Press, 1974. [↑](#footnote-ref-8)
9. Chapman, *Schools as Sorters*, 4; Danziger, *Naming the mind*, 75. [↑](#footnote-ref-9)
10. Starch, Daniel. “Reliability and distribution of grades.” *Science* (1913): 630-636; Starch, Daniel, and Edward C. Elliott. “Reliability of grading work in history.” *The School Review* 21, no. 10 (1913): 676-681; Starch, Daniel, and Edward C. Elliott. “Reliability of grading work in mathematics.” *The School Review* 21, no. 4 (1913): 254-259; Starch, Daniel, and Edward C. Elliott. “Reliability of the grading of high-school work in English.” *The School Review* 20, no. 7 (1912): 442-457. [↑](#footnote-ref-10)
11. Banker, Howard J. “The Significance of Teachers’ Marks.” *The Journal of Educational Research* 16.4 (1927): 271-284, 271. [↑](#footnote-ref-11)
12. Rugg, Harold O. “Teachers' marks and the reconstruction of the marking system.” *The Elementary School Journal* 18, no. 9 (1918): 701-719, 701. [↑](#footnote-ref-12)
13. Rugg, “Teachers’ marks.” [↑](#footnote-ref-13)
14. Id.,, 702, original emphasis. [↑](#footnote-ref-14)
15. Id., 717. [↑](#footnote-ref-15)
16. Steele, A. G. “Training teachers to grade.” *The Pedagogical Seminary* 18, no. 4 (1911): 523-532. [↑](#footnote-ref-16)
17. Rugg developed several; so did Wagner, C.A., 1921. The Construction of a Teacher-Rating Scale. *The Elementary School Journal*, *21*(5), pp.361-366. [↑](#footnote-ref-17)
18. Sharp, L. A. “A study of school marks.” *Peabody Journal of Education* 1, no. 4 (1924): 207-214; Sharp, L. A. “The value of standards in grading examination papers.” *Peabody Journal of Education* 3, no. 1 (1925): 38-45; Byrne, Lee. “A method of equalizing the rating of teachers.” *The Journal of Educational Research* 4, no. 2 (1921): 102-108. [↑](#footnote-ref-18)
19. E.g. ‘The objections to teachers’ judgment as a basis of selection are now familiar through the studies made by Binet and Simon preparatory to the working out of the intelligence tests followed by those of Terman of a similar nature’ (Richards-Nash, Albertine A. “The Psychology of Superior Children.” *The Pedagogical Seminary* 31, no. 3 (1924): 209-246, 212.) [↑](#footnote-ref-19)
20. Arguably the most influential study of psychometry’s founding myths, mistakes and mysticism is Gould, Stephen J. *The Mismeasure of Man*. London: Penguin, 1992. An excellent article (in French) on the differences between Binet, Terman and Thurstone in the statistical choices made to calculate I.Q. is by Olivier Martin. “La mesure en psychologie de Binet à Thurstone, 1900-1930.” *Revue de synthèse* 4, no. 4 (1997): 457-493. Perhaps the most complete exploration of the mathematical and conceptual problems with I.Q. testing from its origins to the 1990s can be found in Fisher, Claude S., Michael Hout, Martin Sanchez Jankowski, Samuel R. Lucas, Ann Swidler & Kim Voss. *Inequality by Design: Cracking the Bell Curve Myth*. Princeton, NJ: Princeton University Press, 1996. [↑](#footnote-ref-20)
21. Brown, *The definition of a profession*, 4. [↑](#footnote-ref-21)
22. Id., 47. [↑](#footnote-ref-22)
23. Hollingworth, Leta S., *Gifted Children: Their Nature and Nurture*. New York: Macmillan, 1926. [↑](#footnote-ref-23)
24. Id., 47. [↑](#footnote-ref-24)
25. Hollingworth, Gifted Children, 50, original emphasis. [↑](#footnote-ref-25)
26. Terman had been elected President of the APA; the previous year, a symposium on intelligence and its measurement had resulted in an influential special issue of *Journal of Educational Psychology*, and the National Intelligence Tests had been presented to the public. Hollingworth’s research on giftedness was beginning to be funded, and Terman had also just recruited researchers for what would become his famous longitudinal studies of gifted subjects. This was also the year of Lewis Terman’s debates with Walter Lippmann and William Bagley (to which we return later). [↑](#footnote-ref-26)
27. Franzen, Raymond. *The Accomplishment Ratio: A Treatment of the Inherited Determinants of Disparity in School Product*. New York City: Teachers College, Columbia University, 1922. [↑](#footnote-ref-27)
28. ‘Methods of determining superiority in school have progressed from the subjective (the teacher’s estimate) to the objective (mental tests)’ (Richards-Nash, “The psychology of superior children”, 241.) It is worth noting that we are here talking about individual intelligence tests, like the Stanford-Binet, rather than group intelligence tests which, while more easily applied to large populations, were not as usable for the purposes of ability grouping. [↑](#footnote-ref-28)
29. Banker, Howard J. “The Significance of Teachers’ Marks.” *The Journal of Educational Research* 16, no. 3 (1927): 159-171, 159. [↑](#footnote-ref-29)
30. Woodrow, Herbert H., 1919. *Brightness and dullness in children*. JB Lippincott Company, 19. [↑](#footnote-ref-30)
31. Terman, Lewis Madison. *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon intelligence scale*. Houghton Mifflin, 1916, 13. [↑](#footnote-ref-31)
32. Terman, Lewis Madison. *The intelligence of school children: How children differ in ability, the use of mental tests in school grading and the proper education of exceptional children*. Mifflin, 1919, 309. [↑](#footnote-ref-32)
33. Tyack, *The one best system*, 59. [↑](#footnote-ref-33)
34. E.g. McAndrew, William. “Women School Teachers.” *Journal of Education* 65, no. 14 (1907): 372-373; Donovan, H. L. “Educating the teacher for the progressive public school.” *Peabody Journal of Education* 9, no. 5 (1932): 266-273 ; Gray, William S., Ralph E. Carter, Leonard V. Koos, and Guy M. Hoyt. “Recruiting Capable Men for the Teaching Profession.” *The Phi Delta Kappan* 4, no. 3 (1922): 4-9. [↑](#footnote-ref-34)
35. Goddard, Henry Herbert. “The gifted child.” *The Journal of Educational Sociology* 6, no. 6 (1933): 354-361, 358. [↑](#footnote-ref-35)
36. Rogers, Karen B. “The Lifelong Productivity of Terman's Original Women Researchers.” Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, March 24-28, 1997). [↑](#footnote-ref-36)
37. Terman, *Schoolchildren*, 299. [↑](#footnote-ref-37)
38. Terman, *Measurement*, 24, 31. Also Hollingworth: teachers ‘do not have in mind a clear idea of the meaning of intelligence. … This results in deviation from the findings of scientific test’ (Gifted Children, 50.) [↑](#footnote-ref-38)
39. Hollingworth, *Gifted Children*, 50. [↑](#footnote-ref-39)
40. Terman, *Measurement*, 23. [↑](#footnote-ref-40)
41. Banker, “The significance of teachers’ marks”, 271. [↑](#footnote-ref-41)
42. Id. [↑](#footnote-ref-42)
43. Terman, *Measurement*, 34. [↑](#footnote-ref-43)
44. Terman, L.M., Dickson, V.E., Sutherland, A.H., Franzen, R.H., Tupper, C.R. and Fernald, G.M., 1923. *Intelligence tests and school reorganization*. World Book Company, 78. This collective book is evocatively introduced by Terman in a chapter called ‘The Problem’ ; the problem being that ‘school reform has lagged behind the advances of psychological science’ (4). [↑](#footnote-ref-44)
45. Hollingworth, Leta., 1942. *Children above 180 IQ: Origin and Development*. London: George G. Harrap, 288. [↑](#footnote-ref-45)
46. Terman, *Measurement*, 30. [↑](#footnote-ref-46)
47. Schwegler, Raymond A. *A Teachers’ Manual for the Use of The Binet-Simon Scale of Intelligence*. Topeka: University of Kansas, 1914, 3. [↑](#footnote-ref-47)
48. Terman, *Schoolchildren*, 291. [↑](#footnote-ref-48)
49. Thus Woodrow, for a typical example: ‘Clearly, the science of education depends upon, and finds its surest foundation in, the science of intelligence’ (*Brightness and Dullness*, 18.) [↑](#footnote-ref-49)
50. Terman, *Measurement*, 35. [↑](#footnote-ref-50)
51. Id., 19. [↑](#footnote-ref-51)
52. Terman, *Schoolchildren*, xiv. [↑](#footnote-ref-52)
53. In Terman, *Schoolchildren*, viii. [↑](#footnote-ref-53)
54. Writing as BRB. “Mental and Physical Age in Relation to School Administration.” *The Journal of Educational Research* (1920): 139-143, 143. [↑](#footnote-ref-54)
55. See Zenderland’s account in *Measuring Minds*; or, more recently, for a case study, Perrillo, Jonna. “Between the School and the Academy: The Struggle to Promote Teacher Research at Columbia University's Lincoln School, 1917–1935." *History of Education Quarterly* 56, no. 1 (2016): 90-114. [↑](#footnote-ref-55)
56. Judd, “The influence”, 297. [↑](#footnote-ref-56)
57. For accounts (from all sides) of Terman’s canny sense of self-promotion and interest in the sales of his tests, see Gould, *The Mismeasure of Man*; Chapman, *Schools as Sorters*, 103; Jolly, Jennifer L. “Historical Perspectives: Pioneering Definitions and Theoretical Positions in the Field of Gifted Education.” *Gifted Child Today* 28, no. 3 (2005): 38-44. [↑](#footnote-ref-57)
58. Terman, Lewis M. “Communications and discussions: Concerning psycho-clinical expertness.” *Journal of Educational Psychology* 5, no. 3 (1914): 164-165, 165. [↑](#footnote-ref-58)
59. Wallin, J. E. “Danger signals in clinical and applied psychology.” *Journal of Educational Psychology* 3, no. 4 (1912): 224. [↑](#footnote-ref-59)
60. Terman, *Schoolchildren*, 38. [↑](#footnote-ref-60)
61. See Reese, William J. *Power and the promise of school reform: Grassroots movements during the progressive era*, London: Teachers College Press, 2002; Cuban, Larry. “Why some reforms last: The case of the kindergarten.” *American Journal of Education* 100, no. 2 (1992): 166-194. [↑](#footnote-ref-61)
62. E.g. Brooks, Samuel S. “Getting teachers to feel the need for standardized tests.” *The Journal of Educational Research* 2, no. 1 (1920): 436-451. [↑](#footnote-ref-62)
63. While too vast a topic to probe in this study, such openly autosuggestive rhetoric may not be unconnected, besides the recent war effort, to the 1920s trend of methods such as Emile Coué’s (Coué, Emile. *Self mastery through conscious autosuggestion*. Malkan Publishing Company, 1922.) [↑](#footnote-ref-63)
64. Bell, J. Carleton. “Educational measurement and the classroom teacher.” *Journal of Educational Psychology* 10, no. 2 (1919): 111-112, 112. [↑](#footnote-ref-64)
65. See Chapman, *Schools as Sorters*, 140-142; Tyack, *The one best system*, 98. [↑](#footnote-ref-65)
66. Raftery, Judith R. “Missing the Mark: Intelligence Testing in Los Angeles Public Schools, 1922-32.” *History of Education Quarterly* 28:1 (1988): 73-93. [↑](#footnote-ref-66)
67. Urban, Wayne J. *Gender, Race, and the National Education Association: Professionalism and Its Limitations*. London: RoutledgeFalmer, 2000. [↑](#footnote-ref-67)
68. Goldstein, *The Teacher Wars*, 90; Lagemann, Ellen Condliffe, *An Elusive Science: The Troubling History of Education Research*. London: University of Chicago Press, 2000. [↑](#footnote-ref-68)
69. Anon. ‘New Era in Schools Urged on Teachers.’ *The New York Times,* Oct. 31, 1931. [↑](#footnote-ref-69)
70. Watson, Goodwin. ‘Intelligence Test Has Limited Scope’. *New York Times* (Oct. 5, 1930). [↑](#footnote-ref-70)
71. Anon. ‘Intelligence Tests Called Dangerous: Dr. McMurry Declares They Bring Education to “a Low Level”.’ *The New York Times* (12 February 1928). [↑](#footnote-ref-71)
72. Anon. ‘Defends Mind Test in Child Training: California Educator Explains Controversial Points as to Its Merits’. *The New York Times* (22 July 1923). [↑](#footnote-ref-72)
73. Goldstein, *The Teacher Wars*, 89. [↑](#footnote-ref-73)
74. Raftery, “Missing the Mark”, 74. [↑](#footnote-ref-74)
75. On the idealism of American psychology, see Herman, Ellen. (1995). *The Romance of American Psychology: Political Culture in the Age of Experts*. London: University of California Press. [↑](#footnote-ref-75)
76. Terman, *Schoolchildren*, 293. [↑](#footnote-ref-76)
77. Precise indications for purchase are given by Schwegler, *A Teacher’s Manual*, 11. [↑](#footnote-ref-77)
78. E.g. ‘There is not a scale or a test now before the public that cannot be given, scored and tabulated by an intelligent teacher with a little careful practice,’ said Bell (“Educational Measurement”, 111); ‘Every effort has been made to remove obscurity in the statement of the tests themselves’, insisted Schwegler (*A Teachers’ Manual*, 12) [↑](#footnote-ref-78)
79. Gould inserts in *The Mismeasure of Man* an advert for the Stanford-Binet scale, extolling its being ‘simple in application’, ‘reliable’; ‘Scoring is unusually simple’ (178). [↑](#footnote-ref-79)
80. Bliss, “How much mental ability?”; Thurstone, L. L. “A cycle-omnibus intelligence test for college students.” *The Journal of Educational Research* 4, no. 4 (1921): 265-278 ; Tubbs, Eston V. “The selection of teachers.” *Peabody Journal of Education* 5, no. 6 (1928): 323-332; more nuanced, Whitney, Frederick Lamson, and Clark Melville Frasier. “The relation of intelligence to student teaching success.” *Peabody Journal of Education* 8, no. 1 (1930): 3-6. [↑](#footnote-ref-80)
81. ‘If the notion has been current, in America, that the schoolmaster or schoolmistress is an oracle and an outstanding paragon of knowledge and ability, that notion must be let slip in the face of the facts’ (Bliss, W. B. “How much mental ability does a teacher need?.” *The Journal of Educational Research* 6, no. 1 (1922): 33-41, 34.) [↑](#footnote-ref-81)
82. Sharp, “The Value of School Marks”, 209. [↑](#footnote-ref-82)
83. See Tyack, *The one best system*, 97. [↑](#footnote-ref-83)
84. Strong, Edward K. *Introductory Psychology for Teachers*. Baltimore: Warwick & York, 1920, xi. [↑](#footnote-ref-84)
85. Superintendent H.B. Wilson, in the introduction to Dickson, Virgil E. *The Use of Mental Tests in School Administration*. Monograph number Four. Berkeley, California: Board of Education, 1922, 6. Lest it be thought that scientists were merely self-congratulating, many such sentences, as in this instance, were written by education professionals. [↑](#footnote-ref-85)
86. In Terman, *Measurement*, xi. [↑](#footnote-ref-86)
87. Woodrow, *Brightness and Dullness*, 39. [↑](#footnote-ref-87)
88. Bell, “Educational Measurement”, 112. [↑](#footnote-ref-88)
89. BRB, “Mental and Physical Age in Relation to School Administration”, 143. [↑](#footnote-ref-89)
90. E.g. Findlay’s introduction to *The Measurement of Intelligence*: ‘Now many teachers will be prepared to admit the value of these tests…’ (In Terman, *Measurement*, vi.) [↑](#footnote-ref-90)
91. Dickson, Virgil E. *Report of the Department of Research*. Oakland, CA: Oakland Public Schools, 1917-1918; Dickson, Virgil E. *The Use of Mental Tests in School Administration*. Monograph number Four. Berkeley, California: Board of Education, 1922.Dickson, Virgil E., and Elise H. Martens. “Training Teachers for Mental Testing in Oakland, California.” *The Journal of Educational Research* 7, no. 2 (1923): 100-108. [↑](#footnote-ref-91)
92. See Chapman, *Schools as Sorters*, 63. [↑](#footnote-ref-92)
93. Dickson, *Report*, 173. [↑](#footnote-ref-93)
94. Id. [↑](#footnote-ref-94)
95. Dickson, “Training Teachers,” 107. [↑](#footnote-ref-95)
96. Id., 100. [↑](#footnote-ref-96)
97. Trabue, M. R. “Some Pitfalls in the Administrative Use of Intelligent Tests.” *The Journal of Educational Research* 6, no. 1 (1922): 1-11. [↑](#footnote-ref-97)
98. Id., 5. [↑](#footnote-ref-98)
99. Terman, *Measurement*, xi; Terman, *Schoolchildren*, xi. [↑](#footnote-ref-99)
100. Terman, *Schoolchildren*, 294. [↑](#footnote-ref-100)
101. Id., 296. [↑](#footnote-ref-101)
102. Whipple, G.M., 1921. The national intelligence tests. *The Journal of Educational Research*, *4*(1), pp.16-31, 22. [↑](#footnote-ref-102)
103. Dickson and Martens, “Training Teachers”, 101-102. [↑](#footnote-ref-103)
104. E.g Otis, Arthur S., and Herbert E. Knollin. “The reliability of the Binet scale and of pedagogical scales.” *The Journal of Educational Research* 4, no. 2 (1921): 121-142. [↑](#footnote-ref-104)
105. Terman, *Schoolchildren*, 298. [↑](#footnote-ref-105)
106. Terman, Lewis M. “Communications and discussions: Concerning psycho-clinical expertness.” *Journal of Educational Psychology* 5, no. 3 (1914): 164, 164. [↑](#footnote-ref-106)
107. Terman, *Schoolchildren*, 299. [↑](#footnote-ref-107)
108. See Keaney, Julia F. “Teaching and Following-Up Supernormal Children in a Small Public School.” *The Journal of Educational Research* 7, no. 2 (1923): 145-148. [↑](#footnote-ref-108)
109. BRB. “Accuracy.” *The Journal of Educational Research* 8, no. 1 (1923): 63-67. [↑](#footnote-ref-109)
110. Pendleton, Charles S. “How to read pupils' written themes.” *Peabody Journal of Education* 1, no. 5 (1924): 272-280, 272. [↑](#footnote-ref-110)
111. See Urban, Wayne. “Organized Teachers and Educational Reform During the Progressive Era: 1890-1920.” *History of Education Quarterly* 16, no. 1 (1976): 35-52 ; Tyack, *The one best system* ; [↑](#footnote-ref-111)
112. Terman, *Intelligence Tests and School Reorganization*, iii-iv. [↑](#footnote-ref-112)
113. Dickson, *Report*, 225. [↑](#footnote-ref-113)
114. Kallom, Arthur W. “Intelligence tests and the classroom teacher.” *The Journal of Educational Research* 5, no. 5 (1922): 389-399, 390.  [↑](#footnote-ref-114)
115. Woodrow, *Brightness and Dullness*, 310. [↑](#footnote-ref-115)
116. Writing as BRB. “The School as a Selective Agency.” *The Journal of Educational Research* (1921): 138-139. [↑](#footnote-ref-116)
117. Id.,139. [↑](#footnote-ref-117)
118. Id. [↑](#footnote-ref-118)
119. Id. [↑](#footnote-ref-119)
120. BRB. “The Next Step after Testing.” *The Journal of Educational Research* (1920): 855-856, 856. [↑](#footnote-ref-120)
121. Hollingworth, Leta S. “Problems of relationship between elementary and secondary schools in the case of highly intelligent pupils.” *The Journal of Educational Sociology* 13, no. 2 (1939): 90-102, 101. [↑](#footnote-ref-121)
122. Dickson, Virgil E. 1917-1918. *Report of the Department of Research*. Oakland, CA: Oakland Public Schools, 231. [↑](#footnote-ref-122)
123. Alexander, Carter. “Presenting educational measurements so as to influence the public favorably.” *The Journal of Educational Research* 3, no. 5 (1921): 345-358, 358. [↑](#footnote-ref-123)
124. For an extensive account of the emergence of such rhetoric in relation to giftedness, see Margolin, Leslie. *Goodness Personified: The Emergence of Gifted Children*. New York: Aldine De Gruyter, 1994. [↑](#footnote-ref-124)
125. Taylor, Howard. “The gifted child and his education.” *NASSP Bulletin* 20, no. 63 (1936): 51-53, 52. [↑](#footnote-ref-125)
126. Goddard, Henry Herbert. “The gifted child.” *The Journal of Educational Sociology* 6, no. 6 (1933): 354-361, 358. [↑](#footnote-ref-126)
127. For a thorough analysis of definitions of democracy in the intelligence-testing movement, see Minton, Henry L. ‘Lewis M. Terman and Mental Testing: In Search of the Democratic Ideal’ in Sokal, Michael (ed). *Psychological Testing and American Society, 1890-1930*. London: Rutgers, 1987, 95-112. [↑](#footnote-ref-127)
128. Donovan, H.L., 1932. Educating the teacher for the progressive public school. *Peabody Journal of Education*, *9*(5), pp.266-273, 273. [↑](#footnote-ref-128)
129. Lagemann, *An Elusive Science*, 22. [↑](#footnote-ref-129)
130. Dewey, John, *The Sources of a Science of Education*. New York: Horace Liveright, 1929, pp.13-14. [↑](#footnote-ref-130)
131. id., 65. [↑](#footnote-ref-131)
132. id., 48. [↑](#footnote-ref-132)
133. Payne, Bruce E. “Ten years of progress in the service rendered by normal schools and teachers colleges 1.” *Peabody Journal of Education* 6, no. 2 (1928): 67-72, 72. [↑](#footnote-ref-133)