



ELSEVIER

Assessing Writing 8 (2002) 5–16

---

---

## ASSESSING WRITING

---

---

# The scope of writing assessment<sup>☆</sup>

Liz Hamp-Lyons\*

*The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong, PR China*

---

*Keywords:* Writing assessment; Historical; Linguistic; Social

---

## 1. Why good writing assessment matters

Does good writing assessment matter? If you have opened these pages and begun to read this article, your personal answer is probably “yes.” But ask 99% of the rest of world and their answer would be “huh”? If we have to tell those people, what do we say?

First, I suppose we say that it matters because so much of it is done by so few to so many, and yet, it remains so unexamined (if you will excuse the pun). Second, it matters because in the twenty-first century the written word remains a principal medium of communication and one of the hopes for the achievement of understanding between peoples everywhere. Following from this is the very important third reason: that access to written language, and to writing in English in particular, remains a ‘good’ that greatly influences access to many, even, most, other ‘goods’ in the twenty-first century world. This makes the assessment of writing an implicitly political act.

There is a fourth reason that is of particular academic interest to readers of *ASW* (and readers are invited to submit responses to this article that suggest further reasons). When writing is assessed, that which is assessed is less well-understood than many other constructs. The questions: What is good writing? How do we know? According to whom? and similar ones, remain intriguing because they

---

<sup>☆</sup> This paper, appearing in my first issue as Editor of *Assessing Writing: An International Journal*, attempts to put the field of writing assessment into an historical, linguistic, and geographical/cultural context, in order to stimulate work in writing assessment from across as broad a set of perspectives as possible.

\* Tel.: +852-2766-7544; fax: +852-2333-6569.

*E-mail address:* eglhl@polyu.edu.hk (L. Hamp-Lyons).

remain unresolved. This journal will make its contribution to exploring these questions.

## 2. A history of writing assessment

### 2.1. *Origins*

It has become common — at least in the US — to think of the direct assessment of writing as a reaction against and result of multiple-choice, ‘objective,’ testing. But in fact essay testing has been around for thousands of years, although it should be better labelled assessment *through* writing than assessment *of* writing. In the Chou period (1111–771 B.C.) writing was one of the “six arts” through which sons of the nobility were prepared for the service of the Imperial Court, using the writing system that had been developed more than a thousand years earlier (and that is basically the same today). In the Eastern Chou period (770–255 B.C.), when Taoism, Confucianism and Mohism flourished, scholars of widely differing perspectives put forward their philosophical and political beliefs in writing. Throughout this thousand-year period, writing skill was rigorously evaluated as a prerequisite for imperial, governmental or feudal service. The establishment of a national university during the Han period (206 B.C.–220 A.D.) formalised the system of written examinations. However, it was during the Sung period (960–1280 A.D.) that a national school system was established and commoners were able to progress in life through education. Education, notably memorisation, rote repetition and written analysis, and literacy became the mark of the successful man (education for women was rare). The spread of block printing during this period made it easier for hopeful candidates to prepare themselves for the exceptionally rigorous examinations, and also opened up the classical system to more people. The wider availability of education placed pressures on the traditional system for selecting officials. A key response to these pressures was ‘*kung*,’ the idea of impartiality (Lee, 1985), which consciously opened up chances for common people from rural areas, as well as the sons of city aristocrats, to compete for official posts. Impartiality in the examination process was ensured through a rigorous, indeed traumatic sequence of increasingly-demanding exams in which candidates and examiners were locked away together. However, in practice these ideals were marred by bribery, cheating and sometimes extreme measures such as tunnelling below exam cubicles to bring in books from outside! (Cleverley, 1985).

It was also during the Sung period that the concept of impartiality was first questioned as far as it related to social justice. The government found that impartiality could not address the need for social, regional and moral justice; while it could (or at least, could aggressively attempt to) create equal possibilities for all candidates, it could not create equal opportunities for all members of the society (Lee, 1985). Shen Kou (who lived from 1028 to 1067) believed this was because it “concentrated only on laying down detailed regulations . . . (and) therefore did not know

what the ‘great fairness’ (*ta-kung*) in the world is, and ignored the fundamentals of governing a state” (quoted in Lee, 1985, p. 205).

## 2.2. *Writing assessment comes to Europe*

In Christian Europe, the concept of written examinations as a means for the democratisation of education developed much later. While the perfection of literacy (usually in Greek or Latin rather than the vernacular) was greatly valued, its practice was usually found in monasteries and seminaries, for until the Renaissance literacy was the prerogative of the clergy and the nobility. DuBois (1970) reports that it was the Jesuit order that pioneered the use of written tests in the West, publishing a statement of writing test procedures in 1599 (op. cit., p. 8). The Jesuits found the highly-formalised examination an ideal method of controlling the teaching in Catholic schools, with instruction monitored by monthly examinations overseen by principals. Spolsky (1996) describes how this convenient form of social control spread throughout the French education system and then across Europe.

Between the Middle Ages and the late nineteenth century, university examinations were typically conducted orally (Rashdall, 1895): this practice continues to the present day in the doctoral *viva voce* exam. In Britain and Europe, university education, open only to a tiny (male) elite, had always consisted primarily of tutor–student dialogue and seminar-style debate and enquiry. Latham (1877) describes how the oral disputation system gradually gave way to written examinations, and the written examinations of the venerable universities of Oxford and Cambridge — most notable in its influence, the Cambridge Tripos — began to wield great symbolic power in defining the ‘man of distinction.’ As British colonial power grew, the demand for literate company clerks and minor government bureaucrats could not be filled entirely from the ranks of the traditional civil service — younger sons of minor nobility. Increasing numbers of middle-class and even some few working class young men sought to join the civil service. In 1833, Macaulay argued to Parliament for the application of the “Chinese principle” in selecting recruits for the Indian Civil Service; he made this argument again 20 years later, and it was passed by the British Parliament in 1853, with the first examinations being held in 1858 (Spolsky, 1996). Written examinations became the simple solution to the need for a relatively quick and dependable way to judge the literacy skills *and* the intelligence of would-be civil servants. Latham (1877, p. 1) complained that examinations were an “encroaching power” that was blurring the distinctions between liberal and technical education, narrowing the range of learning by forcing students to ‘cram’ for exams, and making examinations the master rather than the servant of teaching by exerting control over curriculum. But the criticisms more often levelled against written examinations focused not on their departure from the principles of liberal humanistic education, but on their lack of scientific rigour. First to strike a blow for scientific precision was Edgeworth (1888), who claimed to have discovered a method of “true judgement” of the “intellectual worth” revealed by a written examination text. His proposed

method, which involved the use of several “competent critics” such as literary scholars and textbook authors, whose judgements are then pooled and reported as a mean score, is not so very different from most of the methods commonly used now. His work presaged a host of other attempts to reduce the complexity of the judgement of writing to a task that could be routinised and made objective. Hartog (1910), himself an educational psychologist, sums up the issues in an 1910 entry for the *Encyclopaedia Britannica*:

It can scarcely be doubted that in spite of the powerful objections that have been advanced against [written] examinations, they are, in the view of the majority of English people, an indispensable element in the social organisation of a highly specialised democratic state, which prefers to trust nearly all decisions to committees rather than to individuals. But in view of the extreme importance of the matter, and especially of the evidence that, for some cause or other (which may or may not be the examination system) intellectual interest and initiative seem to diminish in many cases very markedly during school and college life . . . , the whole subject [of written examinations] seems to call for a searching and impartial inquiry. (49)

### 2.3. *Writing assessment comes to the New World*

In the latter part of the nineteenth century and first years of the twentieth, the methodology of the assessment of written work at high school and university in the US had gone through changes and developments similar to those in Britain and often influenced by them (Roach, 1971). As Lunsford (1986) describes, the introduction of written exams in US universities can be traced to Harvard University’s 1873–1874 introduction of a written composition as an entrance examination, replacing the traditional oral examination. The Harvard method entailed “a short composition, correct in spelling, punctuation, grammar and expression, the subject to be taken from such works of standard authors as shall be announced from time to time” (quoted in Applebee, 1974, p. 30). The Harvard approach was opposed by many, most notably Fred Newton Scott of the University of Michigan, and this opposition led to the formation of the US’s National Council of Teachers of English in 1911 (Hook, 1979). Among many texts illuminating what US education lost in the move away from the nineteenth century educational system of oral disputation *The Education of Henry Adams* is perhaps the best known (Adams, 1961).

The rapid increase in the numbers of test candidates led to a search for more expedient, quicker ways of judging written texts, and therefore to the attention of the new field of educational measurement. Lunsford (1986) characterises this period:

Where reading, writing and speaking had once been combined in the pursuit of a student’s own academic and social goals, writing was now separated from the other communicative arts. The direct result of this separation was a dramatic loss of purpose: writing became not primarily a means of influencing important public affairs but merely a way to demonstrate proficiency. Divorced from its original purpose in rhetorical instruction, writing shifted its focus from discovering and

sharing knowledge to being able to produce a “correct” essay on demand; lost the theoretical framework that related language, action and belief; and became increasingly preoccupied with standards of usage, a tendency that grew, by the turn of the century, into a virtual cult of correctness. (p. 6)

Further in the US then the move from a liberal humanistic to a behavioural scientific view of writing ability went further than in Europe. Educational researchers such as Paterson (1925), Miller (1926), and Odell (1928) drew on work on the measurement of human cognitive ability by psychologists such as Thorndike (1904), Yerkes, Bridges, and Hardwick (1915) and Yerkes (1921) in developing the theory and model of what Paterson (op. cit.) began to call “new-type” examinations.

A key element in the solution they proposed to “the tension between the unmistakably effective and awesome power of the examination and the manifest uncertainty of its scores” (Spolsky, 1996) involved separating the content of a text from its form. Once this was done, writing as the expression and creation of meaning became lost from view. Over the period 1920–1960, a view of literacy as a mechanical ability composed of several discrete skills that could be measured separately — and for practicality and financial reasons, should be — became predominant in the US. Despite concerns raised by some powerful forces, such as the College Board (Thomas, 1931) that the form of the examinations by which a student is judged will have a strong influence on what s/he is taught and learns, and that the “new-type” indirect examinations of writing might threaten the ‘civilising influence’ of education, by the late 1950s and 1960s most educational assessment research was focused on so-called ‘objective’ testing, and the assessment of writing was caught up in this behaviourist paradigm. Yancey (1999) reports that a 1952 survey of US undergraduate colleges/universities by Sasser found that, not only did the overwhelming majority of institutions place students into writing courses with standardised tests: nearly half of these institutions also used a standardised test as, or as part of, the final grade from the writing course!

#### 2.4. *Slightly different histories*

But in the UK there were stronger objections to such a reductionist view of what writing is and how it can be judged. The work of Wiseman (1949, 1956) is best known in this regard. Wiseman argued that more attention should be given to validity, and stressed what he called the “backwash effect,” the tendency to de-emphasise actual writing as an ability deserving to be evaluated with the resulting decrease in attention to the teaching of writing. Wiseman’s *backwash* concept — more usually called *washback* these days — is still a powerful and important one in British educational assessment. Wiseman and his colleagues also developed and implemented the so-called ‘Devon’ method of multiple marking of compositions, which has been used in various forms in the UK since the 1940s.

Wiseman’s work was at the school level, not college level. Perhaps because the British educational system was at that time (and perhaps still is) considerably

more elitist than the American, with only 5% of 18-year-olds going on to higher education, there was no equivalent of “freshman writing” in UK colleges and universities. The ability to write was taken for granted. At the university level then, Britain could escape the concerns about writing assessment that were so painfully felt in the US. At the high school level, the traditional “exam boards” such as the University of Cambridge Local Examinations Syndicate, the Oxford Delegacy, and others, which had been using much the same practices since the turn of the twentieth century, were in general fairly content to continue using written exams.

### 3. The modern period

Huot (1990) locates the mid-1960s as the beginning of a time of change in writing assessment: this date coincides with the work of Diederich, French, and Carlton (1961), spurring serious research into improving direct writing assessment. Huot (op. cit.) also says: “it appears that the political and economic climate for the direct testing of writing was not suitable until the middle of the 1970s” (pp. 237–238), and sees Cooper and Odell’s (1977) anthology as the watershed publication. The Cooper and Odell anthology describes the three main approaches that are still current today, a quarter of a century later: analytic scoring, primary trait scoring, and holistic or impression marking. This article does not look at the history of the modern direct assessment movement. Valuable overviews are found in Greenberg, Wiener, and Donovan (1986) and White, Lutz, and Kamusikiri (1996) and for non-native speakers, Hamp-Lyons (1991). Bringing us closer to the present, Yancey (1999) identifies three ‘waves’ of writing assessment: the first wave (1950–1970), when writing assessment was done through ‘objective’ testing; the second wave (1970–1986), when holistic scoring of timed essays was the preferred practice; and the third wave (1986–present) with its interest in portfolio assessment (p. 484). Hamp-Lyons (2001) argues that the first two ‘generations’ of writing assessment were (as more or less illustrated in the preceding parts of this chapter) the opposite of Yancey’s characterisation: direct assessment of written texts came first, and multiple-choice testing second. Nevertheless, there is little disagreement that the last 15 years of the twentieth century turned the attention of writing assessment specialists and many other educators to portfolios as a fruitful form of assessment.

#### 3.1. *Portfolio-based writing assessment*

Hamp-Lyons and Condon (2000) explain that the impetus for portfolios for writing assessment came not from assessment specialists but from teachers:

... just as the battle to establish the legitimacy of direct assessment of writing seemed to be ending in victory, teachers of writing were becoming increasingly

dissatisfied with direct tests of writing and with holistic scoring. . . . although such a direct sample is a far superior instrument than the indirect test, teachers increasingly saw it as a context-poor assessment and began looking for an even better instrument. (p. 12)

The dissatisfaction had (and has) two main causes: lack of authenticity/validity [what is sometimes called in the field of educational measurement *ecological validity*]; and washback [the effects of this form of testing on teaching]. Experimentation with portfolio assessment of writing began as a search for a way to overcome these criticisms (Belanoff & Elbow, 1986; Elbow & Belanoff, 1991) and took off with astonishing rapidity. As time has gone, however, portfolios have ceased to be seen as the panacea for all issues in evaluating writing ability, and the need for careful and intensive programmes of familiarisation, training, and validation such as described and recommended by Hamp-Lyons and Condon (op. cit.) has been accepted by the profession. The positive benefits of a portfolio-based assessment on the teaching community as described by Belanoff and Elbow (1986) and contributors to the Belanoff and Dickson (1991) anthology have proved themselves repeatedly: but the use of portfolios for formal, high stakes assessment is still problematic.

### 3.2. *Computer-based writing tests*

A very new development for writing assessment is the possibility of assessing writing online. At least in theory, online writing assessment allows the development of tests that fit the levels, purposes and needs of test-takers, as well as of future employers, admissions officers, etc. It allows the design of writing tasks that can be built out of multiple elements previously identified, ensuring authenticity and task variety. A really “smart” computer-based writing assessment system would enable each writer to build a portfolio of writing and select from it according to a particular purpose. A computer-based writing assessment would enable multiple pathways for writers through the many pitfalls of tests. The possibilities for computer-based writing assessment are not limited by the computer software but in what ‘the system’ [educational, financial, political] will allow.

### 3.3. *Second-language and ‘non-traditional’ writers*

One of the strongest characteristics of the modern period, in writing assessment as in everything else in writing instruction, is the appearance in the education systems of English Native Language (ENL) countries of very large numbers of second-language and other ‘non-traditional’ learners. Second-language learners are tremendously varied in language background and degree of cultural integration, as well as socio-economic status, personality, learning style, as well as all the other factors which apply equally to L1 learners. If a writing assessment is to be humanistically, as well as psychometrically, defensible, all of these factors should be accounted for. As the work of Basham and Kwachka (1991) with the Athabaskan

people of Alaska, and of Ballard and Clanchy (1991) with Asian learners in Australia showed, we do not yet possess sufficient knowledge of culturally determined writing behaviours to be able to teach students what to change in their writing in order to conform to expectations, should they wish to do so. Although, it is now fairly well accepted that written text production is in part culturally determined (Connor, 1996; Connor & Kaplan, 1987), there is still far too little research for it to be possible to make assertions about the “usual” cultural patterns and/or problems of writers from any particular background (Hamp-Lyons & Zhang, 2001). The current trend in the field of applied linguistics is for serious questions to be raised about the dominance of English in the academic world, and for “other Englishes” such as Indian English, Singapore English and Jamaican English to be given greater respect. In the US alone there are millions of schoolchildren and college students who are non-native speakers/writers of English, and the design of appropriate instruction and assessment for them is a key concern to school districts, college support programs and professional organisations.

But as many American educators have noted, among them Farr (1993) and Kamusiriki (1996), similar difficulties arise for users of African American English (Black English Vernacular, or Ebonics), who are native speakers of English — just a slightly different English. The genres privileged on most direct writing assessments demand what Farr (op. cit.) calls “essayist literacy” and penalise those who write otherwise. Epps (1985) takes an extreme position when she says: “A wholesale slaughter of Afro-Americans is taking place in this country (the US) every day! . . . The minds of black students are being robbed and mugged on a daily basis because they are not being taught to read and write so that they can determine the course of their own lives” (p. 154). Similar concerns for literacy rights, particularly the right to academic literacy, are being raised in the UK, Canada, Australia and New Zealand. As Foucault (1980) says: “What makes power hold good, what makes it accepted, is simply the fact that it doesn’t only weigh on us as a force that says no, but that it traverses and produces things, it induces pleasure, forms knowledge, produces discourse” (p. 119). The power of discourses must be shared with learners of all backgrounds; and the discourses of education at once hold power and promise power. But not only must powerful discourses be opened to less powerful groups: what Foucault has called “subjugated knowledges” (op. cit., p. 82) should be opened up and given credence so that they become accessible to users of more powerful, mainstream discourses: this is the motivation for critical discourse analysis (see, for example, Fairclough, 1992).

#### **4. Where next?**

In a recent paper (Hamp-Lyons, 2001), I argued that while the third ‘generation’ of writing assessment still has a long way to go before all problems are resolved, there is a need to look forward to the fourth generation. I suggested (pp. 120–125) that a fourth generation will need to be technological, humanistic, political, and

ethical. The power of computing and the World Wide Web must be harnessed to empower not only large test agencies, but more importantly test-takers, raters and educators. At the same time, resisting the potential of computer-based assessment to dehumanise and automatize the testing act, the needs of stakeholders must be the focus of a humanistic ‘turn’ in attention to writing assessment. The huge variation in stakeholders — particularly learners — and their backgrounds and needs must be acknowledged in testing solutions. This fourth generation must be both humanistic and technological, drawing on advances both in computer applications and in our increasing understanding of writing assessment as a complex of processes in which multiple authors and readers are involved and revealed (see, for example, Lumley, 2000).

But further, the political nature of all assessment must be realised, and we must all come to terms with the fact that what we do when we design or administer a [writing] test, when we score it, when we take and utilise test scores, is to participate in a form of social engineering that is at once beneficial and dangerous. As we settle into the educational climate and values of the twenty-first century, it seems that all around the world those who administer educational systems, and those responsible for the financial decisions which determine the fate of education, are increasingly turning to tests and scores to underpin and to justify their decisions, and that this turn to tests as planning and policy tools will increase still further. The recent announcement by the Bush (Jnr) government in the US of the ‘No Child Left Behind’ policy is a dramatic example of a good idea gone bad with an unjustified and unjustifiable dependence on and faith in tests.

But the United States is by no means alone, and similar trends in education have been seen since the days of the Thatcher government in the UK, and the trend is toward government “benchmarks,” for example, 85% of 11-year-old pupils at Level 4 (of a 8 level scale) in English by 2004, accompanied by centrally produced standardised tests to track progress toward the benchmarks. Similarly in Canada, many educational jurisdictions have over the past decade implemented tests to monitor the progress of students in school systems, whereas the previous tradition had been to only administer large-scale tests as students completed their final year of secondary schooling. In Australia too, the past 15 years have shown tremendous growth in tests and scales, and systems for measuring development in writing and literacy ability. In New Zealand, the situation is different: the previous government had plans for nation-wide testing of primary school students but these were scrapped by the present centre-left coalition: only in the case of international students, skilled and business migrants, and applicants to professional registration bodies is large scale testing of writing ability carried out.

As someone who has made a career out of writing assessment research and practice, I cannot and do not oppose writing assessments. But I do oppose the unnecessary use of writing tests where other more appropriate and less intrusive tools may do the job. I do support the development and use of high-quality alternative forms of writing assessment such as portfolios and authentic work-sampling. I do support the efforts of teachers to use response to writing, peer feedback,

self-reflection and self-assessment, and all sound, creative paths open to them to place writing assessment in support of teaching. When writing tests are necessary — as they are — I support the development, validation and use of high-quality tests within fully-accountable systems. So too do all readers of *Assessing Writing*.

This takes me to the last of the four dimensions of my fourth generation in writing assessment: the ethical dimension. In educational measurement, and in second-language testing, a focus on issues of ethicality of testing practices has been a major research development in the past 7/8 years (see, for example, the special issue of *Language Testing on Ethics in Language Testing*, 14 (3) 1997). In writing assessment, however, that dimension has been relatively unexplored (but see [Greenberg, 1998](#)). The growth in frequency of test use and the rising stakes attached to performance on tests is making us all aware of the power of tests to shape the lives of the people we teach and assess. The ethical dilemmas and challenges we face in balancing society's need for assessments with our determination to do our best for learners are very great. Accepting a shared responsibility for the impact of writing assessment practices will put consideration of our own ethical behaviour at the top of our agenda.

## 5. A future of writing assessment — in this journal

The need for *Assessing Writing: An International Journal* is, therefore, as great as or greater than it was 8 years ago when Kathi Yancey and Brian Huot, with the encouragement of us all, started *Assessing Writing*. Since then our world has got both larger and more complex, and smaller and more textured. We have solved problems, discovered others, and certainly made some mistakes. Engaging in scholarly and practical discourse through the pages of this journal, we will improve ourselves and the work we do.

## References

- Adams, H. (1961). *The education of Henry Adams*. Boston: Houghton Mifflin.
- Applebee, A. (1974). *Tradition and reform in the teaching of English: A history*. Urbana, IL: NCTE.
- Ballard, & Clanchy. (1991). Assessment by misconception: Cultural influences and intellectual traditions. In: L. Hamp-Lyons (Ed.), *Assessing second-language writing in academic contexts* (pp. 19–36). Norwood, NJ: Ablex.
- Basham, & Kwachka. (1991). Reading the world differently: A cross-cultural approach to writing assessment. In: L. Hamp-Lyons (Ed.), *Assessing second-language writing in academic contexts* (pp. 37–50). Norwood, NJ: Ablex.
- Belanoff, P., & Dickson, M. (Eds.). (1991). *Portfolios: Process and product*. Portsmouth, NH: Heinemann/Boynton Cook.
- Belanoff, P., & Elbow, P. (1986). Using portfolios to increase collaboration and community in a writing program. *Writing Program Administration*, 9 (3), 27–40.
- Cleverley, J. (1985). *The schooling of China*. London: Allen & Unwin.
- Connor, U. (1996). *Contrastive rhetoric*. Cambridge: Cambridge University Press.

- Connor, U., & Kaplan, R. (1987). *Writing across languages: Analyses of L2 texts*. Reading, MA: Addison-Wesley.
- Cooper, C. R., & Odell, L. (1977). *Evaluating writing: Describing, measuring, judging*. Urbana, IL: NCTE.
- Diederich, P. B., French, J. W., & Fr Carlton, S. T. (1961). *Factors in judgements of writing ability*. (*Research Bulletin 61-15*). Princeton, NJ: Educational Testing Service.
- DuBois, F. (1970). *A history of psychological testing*. Boston: Allyn & Bacon.
- Edgeworth, F. Y. (1888). The statistics of examinations. *Journal of the Royal Statistical Society*, 51, 599–635.
- Elbow, P., & Belanoff, P. (1991). SUNY: Portfolio-based evaluation program. In: P. Connelly & T. Vilaridi (Eds.), *New methods in college writing programs* (pp. 96–105). New York: Modern Language Association.
- Epps, J. (1985). Killing them softly: Why Willie can't write. In: C. K. Brooks (Ed.), *Tapping potential: English and language arts for the black learner* (pp. 154–158). Urbana, IL: NCTE.
- Fairclough, N. (Ed.). (1992). *Critical language awareness*. London: Longman.
- Farr, M. (1993). Essayist literacy and other verbal performances. *Written Communication*, 10, 4–38.
- Foucault, M. (1980). *Power/knowledge: Selected interviews 1972–1977*. Ed. Colin Gordon. New York: Pantheon.
- Greenberg, K. L. (1998). Grading, evaluating, assessing: Power and politics in college composition. *College Composition and Communication*, 49 (2), 275–284.
- Hamp-Lyons, L. (Ed.). (1991). *Assessing academic writing in a second language*. Norwood, NJ: Ablex.
- Hamp-Lyons, L. (2001). Fourth generation writing assessment. In: T. Silva & P. K. Matsuda (Eds.), *On second-language writing* (pp. 117–128). Mahwah, NJ: Lawrence Erlbaum.
- Hamp-Lyons, L., & Zhang, W.-X. (2001). World Englishes: Issues in and from academic writing. In: J. Flowerdew & M. Peacock (Eds.), *Research perspectives on English for academic purposes* (pp. 101–116). Cambridge: Cambridge University Press.
- Hartog, S. P. (1910). Examinations. *Encyclopaedia Britannica, 1910 Edn*. Cambridge: Encyclopaedia Britannica, Inc.
- Hook, J. N. (1979). *A long way together*. Urbana, IL: NCTE.
- Huot, B. (1990). The literature of direct writing assessment: Major concerns and prevailing trends. *Review of Educational Research*, 60 (2), 237–263.
- Kamusiriki, S. (1996). African American English and writing assessment: An Afrocentric approach. In: E. M. White, W. D. Lutz, & S. Kamusikiri (Eds.), *Assessment of writing: Politics, policies, practices* (pp. 187–203). New York: Modern Language Association.
- Latham, H. (1877). *On the action of examinations considered as a means of selection*. Cambridge: Deighton, Bell & Co.
- Lee, T. H. C. (1985). *Government education and examinations in Sung China*. Hong Kong: Chinese University Press.
- Lumley, T. J. N. (2000). *The process of the assessment of writing performance: The rater's perspective*. Unpublished Ph.D. Dissertation, University of Melbourne.
- Lunsford, A. (1986). The past — and future — of writing assessment. In: K. L. Greenberg, H. S. Wiener, & R. A. Donovan (Eds.), *Writing assessment: Issues and strategies* (pp. 1–12). White Plains, NY: Longman.
- Miller, G. F. (1926). *Objective tests in high school subjects*. Norman, OK: mimeo.
- Odell, C. (1928). *Traditional examinations and new-type tests*. Englewood Cliffs, NJ: Prentice-Hall.
- Paterson, D. (1925). *Preparation and use of new-type examinations: A manual for teachers*. Orlando, FL: Harcourt Brace Jovanovich.
- Rashdall, H. (1895). *The universities of Europe in the Middle Ages*. London: Deighton, Bell & Co.
- Roach, J. (1971). *Public examinations in England 1850–1900*. Cambridge: CUP.
- Spolsky, B. (1996). The examination–classroom backwash cycle: Some historical cases. In: D. Nunan, R. Berry, & V. Berry (Eds.), *Bringing about change in language education* (pp. 55–66). Hong Kong: Department of Curriculum Studies, The University of Hong Kong.

- Thomas, C. S. (1931). *Examining the examinations in English: Report to the college entrance examination board*. Cambridge, MA: Harvard University Press.
- Thorndike, E. L. (1904). *An introduction to the theory of mental and social measurement*. New York: The Science Press.
- Wiseman, S. (1956). Symposium on the use of essays on selection at 11+. *British Journal of Educational Psychology*, 26 (3), 172–179.
- Wiseman, S. (1949). The marking of English compositions in grammar school selection. *British Journal of Educational Psychology*, XIX (3), 200–209.
- Yancey, K. B. (1999). Looking back as we look forward: Historicizing writing assessment. *College Composition and Communication*, 50 (3), 483–503.
- Yerkes, R. M. (1921). *Psychological examining in the United States army*. Washington: Government Printing Office.
- Yerkes, R. M., Bridges, J. W., & Hardwick, R. S. (1915). *A point scale for measuring mental ability*. Baltimore, MD: Warwick & York, Inc.