Lang. Teach. (2009), **42:1**, 95–107 © Cambridge University Press
doi:10.1017/S0261444808005430

Research Timeline

Assessing academic writing in foreign and second languages

Alister Cumming University of Toronto, Canada
acumming@oise.utoronto.ca

Academic writing and education in foreign and second languages both have lengthy histories. Their histories have diverged but also intersected. Matsuda (2005), for example, described the convergence in policies that led to a distinctive discipline of 'L2 writing' in higher education in the U.S.A. during the latter part of the 20th century. Assessment is a key point of their mutual intersection. Approaches to language testing that involve writing for academic purposes – or conversely, writing assessments that involve second or foreign languages – have appeared at certain times, supported by particular policies, rationales, and modes of inquiry or reasoning. The present timeline focuses on the development and elaboration of these rationales and research about them.

Theearly history linking writing assessment and language education predates existing documentation. Nonetheless, it is evident that, long before any explicit surviving records, many conventions persisting today in ordinary practices of writing assessment in second or foreign languages were already well established in education and other institutions (Kelly 1969; Spolsky 1995), including composing or translating set texts, a teacher's or examiner's evaluation of the quality of that performance, and students' subsequent revisions of their texts.

Since the early 20th century, many progressions and differentiations have appeared. My selection of notable publications below aims to represent these historical trends. I observe five recurring themes, each of which is marked by capital letters A–E in the final-column annotations to each publication:

A Broadening of the contexts to which assessment has been applied.
B Concerns for the validity of assessments, particularly scoring methods and the evidence and arguments to support inferences made from writing assessments.
C Congruence and conflict between the purposes of assessment and of education in general.
D Debates over appropriate or variable standards of writing quality.
E An expanding repertoire of research methods and technologies to improve and validate L2 writing tests.

I constructed this timeline and its components by consulting key sources (cited below) and my previous searches, notably in Cumming (1997, 2008) and Leki, Cumming & Silva (2008). Other, more extensive, historical studies were also helpful, particularly Spolsky (1995), but

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

also Kelly (1969), Barnwell (1996), Hamp-Lyons (2002), and Weir & Milanovic (2003). My citation of sources should be consistent with these previous analyses, though my selection and interpretations are personal. Undoubtedly a bias appears towards the literature in English and the educational contexts in North America and parts of Europe or Asia with which I am most familiar. However, I hope that the values I have expressed do not, for these reasons, impose an unduly hierarchical viewpoint on situations that must, around the world, necessarily have their own local realizations and priorities (cf. Canagarajah 2005).

References

Barnwell, D. (1996). *A history of foreign language testing in the United States: From its beginning to the present*. Tempe, AZ: Bilingual Press.

Canagarajah, S. (ed.) (2005). *Reclaiming the local in language policy and practice*. Mahwah, NJ: Lawrence Erlbaum.

Cumming, A. (1997). The testing of second-language writing. In C. Clapham (volume ed.) & D. Corson (series ed.), *Encyclopedia of language and education* (vol. 7): *Language assessment*. Dordrecht: Kluwer, 51–63.

Cumming, A. (2008). Assessing oral and literate abilities. In E. Shohamy (volume ed.) & N. Hornberger (series ed.), *Encyclopedia of language and education* (2nd edn.) (vol. 7): *Language testing and assessment*. New York: Springer, 3–18.

Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing* 8.1, 5–16.

Kelly, L. G. (1969). *25 centuries of language teaching: 500 B.C. – 1969*. Rowley, MA: Newbury House.

Leki, I., A. Cumming & T. Silva (2008). *A synthesis of research on writing in a second language in English*. London: Routledge.

Matsuda, P. (2005). Historical inquiry in second language writing. In P. Matsuda & T. Silva (eds.), *Second language writing research: Perspectives on the process of knowledge construction*. Mahwah, NJ: Lawrence Erlbaum, 33–46.

Spolsky, B. (1995). *Measured words: The development of objective language testing*. Oxford: Oxford University Press.

Weir C. & M. Milanovic (eds.) (2003). *Continuity and innovation: Revising the Cambridge Proficiency in English Examination 1913–2002*. Cambridge: Cambridge University Press.

ALISTER CUMMING is Professor and Head of the Modern Language Centre at the Ontario Institute for Studies in Education, University of Toronto. He conducts research and teaches courses on writing in second languages, language assessment, and research methods. In addition to many articles in journals, book chapters, and research reports, his recent publications include *Goals for academic writing* (John Benjamins, 2006) and, with Ilona Leki and Tony Silva, *A synthesis of research on second language writing in English* (Routledge, 2008). He currently chairs the TOEFL Committee of Examiners at Educational Testing Service and serves as Executive Director of *Language Learning*, a journal he edited in the 1990s.

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
Appx. 606–620	Cheng, L. (2008). The key to success: English language testing in China. *Language Testing* 25.1, 15–37. Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing* 8.1, 5–16. Spolsky, B. (1995). *Measured words: The development of objective language testing*. Oxford: Oxford University Press.	**Cheng**, **Hamp-Lyons** and **Spolsky** attributed the earliest recorded functions of testing to the establishment of university then school systems in China. From the Han through to Sui and Tang Dynasties, written examinations were used to select officials for imperial or feudal service on the basis of merit and knowledge rather than patronage. Selection by merit and with regulation by institutional authority established a fundamental role for written assessments and implicit requirements for their validity.	**A** **B**
1600s–1700s	Foucault, M. (1979). *Discipline and punish: The birth of the prison*. New York: Vinavtage.	**Foucault** argued that in the 17th century Catholic then secular schools in France, and in other European nations, established regular written examinations as a method of normalizing school routines, creating self-regulatory discipline among students, formally archiving knowledge, and thus regulating social control of power. Foucault's argument assumes that assessment is integral to formal education and that the expansion and value of mass education promoted 'discipline', through the self-controlling rigor of preparations for and the performance of examinations among select members of increasingly hierarchically-organized societies. In Elizabethan England, written exams replaced oral exams as students' abilities in spoken Latin decreased.	**C** **A**

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
1627	Brinsley, J. (1927). *Ludus literarius or the grammar schools*. London: Thomas Mann. Kelly, L.G. (1969). *25 centuries of language teaching: 500 B.C. – 1969*. Rowley, MA: Newbury House.	**Brinsley**'s prescriptions, among others quoted and summarized by **Kelly** (pp. 166–180), make clear the existence of well established pedagogical conventions, probably dating back to Roman and Greek rhetorical traditions, of teaching written composition and the translation of classical texts, teachers' correction of students' errors in grammar and style, and students' subsequent revisions of their writing to make corrections. These conventions show the long-standing relations between writing assessment and language learning in education as well as ongoing debates over uses of assessment to maintain standards of writing quality through valid assessments.	**C** **D**, **B**
1877	Latham, H. (1877). *On the action of examinations considered as a means of selection*. Cambridge: Deighton, Bell.	Among others, **Latham** voiced arguments for the value of written examinations (over oral disputations) and traced their historical development and uses at British universities and for the selection – on the basis of merit rather than privilege – of employees for the civil and colonial services. This text exemplifies the expanding scope of written examinations within European higher education and their spread to colonial institutions internationally. Arguments are also expressed for the validity of tests over other, non-systematic methods of selecting personnel for employment or students for higher education. As with the previous three entries, these early rationales for writing assessments established in L1 contexts and institutions were later adopted and applied to policies for assessments in L2 education for purposes such as admissions to universities, selection for employment, and certification of occupational credentials.	**A** **B** & **D**

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

1912	Thorndike, E. I. (1912). *Education: A first book*. New York: Macmillan.	**Thorndike** articulated principles for psychological measurement and the design of standardized tests, including empirically-derived scales for assessing compositions and handwriting. Along with other founders of psychometrics, Thorndike helped to establish methods for the design, data collection, and statistical analysis of evidence to evaluate the validity of tests and their uses in education and other contexts.	**B**, **C** & **E**
1913	Weir, C. & M. Milanovic (eds.) (2003). *Continuity and innovation: Revising the Cambridge Proficiency in English Examination 1913–2002*. Cambridge: Cambridge University Press.	In 1913, the first Cambridge Certificate of Proficiency in English exams appeared explicitly for foreigners, including topics for two-hour English written essays. **Weir & Milanovic** (pp. 2–3) document how the initial topics and format of this exam were preserved from the exam for British examinees, but the exam was applied to a new population entering British universities from its colonies. Broader topics for writing appeared in the exams in subsequent years.	**A**
1929	Henmon, V. (ed.) (1929). *Achievement tests in the modern foreign languages, prepared for the modern foreign language study and the Canadian Committee on Modern Languages*. New York: Macmillan.	In a major effort to survey and evaluate the effectiveness of FL teaching and learning in the U.S.A. and Canada, **Henmon** and colleagues devised detailed scales for ranking and scoring the qualities of written compositions in FLs, working from empirical analyses of language samples, rather than explicit theoretical bases, establishing and applying research and testing methods akin to those that have prevailed until now. Their study resulted, however, in the realization that education in FLs could produce only limited achievements during schooling; this led to a policy emphasis on comprehension in North America, and the subsequent neglect for many decades of oral or written production in instruction or assessment in FLs.	**A**, **B**, **C**, **D** & **E**

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
1961	French, J. W., P. Diederich & S. Carlton. (1961). *Factors in the judgement of written composition*. Princeton, NJ: Educational Testing Service.	In this landmark, in-depth investigation of the factors that constitute effective writing, **French et al.** studied how various groups of raters, drawn from diverse professions, evaluated samples of L1English written compositions. They established the salience of content, development, organization, style, wording, and mechanics as fundamental bases on which to base scoring scales – a conceptualization that has persisted in L1 writing assessments and was later adopted in L2 writing assessments.	**A**, **B**, **D** & **E**
1961	Lado, R. (1961). *Language testing: The construction and use of foreign language tests*. London: Longman.	**Lado** influentially advocated objective methods of using discrete test items to evaluate knowledge of FLs but also the direct, holistic scoring of short, written compositions. His methodological prescriptions informed the rationale and design of tests in many languages, but the prevailing emphasis at the time on audio-lingual or grammar-translation teaching methods led to tests that emphasized grammar and vocabulary, and reading and listening abilities rather than attention to writing abilities.	**A**, **B** & **E**
1961	Carroll, J. B. (1961). Fundamental considerations in testing for English language proficiency of foreign students. In *Testing the English proficiency of foreign students*. Washington, DC: Center for Applied Linguistics, 30–40.	**Carroll** outlined a grid of aspects of language (orthography or phonology, morphology, syntax, lexis) and skills that constituted interrelated factors of language proficiency to be assessed in an 'integrative' (rather than discrete point) manner. This model was utilized in CARROLL's (1975) international surveys and his theories of skill learning. It was also proposed as an initial basis for the TOEFL, but the writing component was not adopted, seemingly for reasons of feasibility, reliability, and expense.	**A**, **B** & **E**

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

1975	Carroll, J. B. (1975). *The teaching of French as a foreign language in eight countries*. New York: John Wiley & Sons. Lewis, G. & C. Massad (1975). *The teaching of English as a foreign language in ten countries*. New York: John Wiley & Sons.	**Carroll**'s and **Lewis & Massad**'s international, comparative surveys established models for assessing FL proficiency – including writing in FLs, through tests of grammar, verb formation, and directed compositions (specifying key words and themes), and scoring for both the quantity and quality of writing – that were adopted widely around the world.	**A**, **B** & **E**
1975	Diederich, P. B. (1975). *Measuring growth in English*. Urbana, IL: National Council of Teachers of English.	**Diederich** produced an influential synthesis of principles and methods for evaluating English L1 writing, consolidating methods for assessment, research, and educational policies.	**B** & **D**
1977	Davies, A. (1977). The construction of language tests. In J. P. B. Allen & A. Davies (eds.), *Testing and experimental methods* (vol. 4): *Edinburgh course in applied linguistics*, 38–104.	In an influential book defining practices and principles for applied linguists internationally, **Davies** described a repertoire of methods for the design of language tests, including methods for writing assessment that involved sentence completion and vocabulary tasks, controlled summaries or framework essays, and free composition and letters.	**A**, **B** & **E**
1977	Qian, D. (2008). English language assessment in Hong Kong: A survey of practices, developments and issues. *Language Testing* 25, 85–110.	**Qian** documented how the Hong Kong Examination Authority was established in 1977 to administer written examinations in English equivalent to O and A level tests in the U.K. This is but one instance of the institutionalization of writing assessments in colonial, post-colonial, and economically developing nations, producing controversies over standards of language and writing quality in reference to local and international norms.	**D** & **E**

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
1979	Kroll, B. (1979). A survey of writing needs of foreign and American college freshmen. *English Language Teaching Journal* 33, 219–227.	**Kroll**, among others, produced research involving needs analysis surveys to determine the types of writing tasks and abilities required on composition tests in L2s. At the same time, such research acknowledged expansions in the populations of adults seeking higher education in English and other international languages at universities.	**A** & **E**
1980	Perkins, K. (1980). Using objective methods of attained writing proficiency to discriminate among holistic evaluations. *TESOL Quarterly* 14, 61–69.	**Perkins** reviewed a variety of objective methods for scoring written compositions, such as uses of T-units, which encouraged the use of detailed text analyses of written compositions as a basis to validate holistic scoring methods.	**B** & **E**
1980	Canale, M. & M. Swain (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics* 1, 1–47.	**Canale & Swain** articulated an argument for the realistic, direct assessment of writing and speaking in FLs, through performance of communication tasks, as evidence of key dimensions of learners' communicative competence. Principles of communicative language teaching required compatible methods of assessing learners' communicative performance.	**B**, **C** & **E**
1981	Jacobs, H., S. Zinkgraf, D. Wormuth, V. Hartfiel & J. Hughey (1981). *Testing ESL composition: A practical approach*. Rowley, MA: Newbury House.	**Jacobs et al.** produced a widely used guidebook for evaluating ESL compositions by weighted, analytic scoring. The assessment scheme drew on growing interests in L1 writing instruction and synthesized established methods to evaluate L1 writing (e.g. DIEDERICH 1975) but capitalized on L2 educators' concerns to see evidence in students' writing quality from pedagogical emphasis on the 'processes of composing'.	**A**, **B**, **D** & **E**
1984	Homburg, T. (1984). Holistic evaluation of ESL compositions: Can it be validated objectively? *TESOL Quarterly* 18, 87–106.	**Homburg** produced data and speculated on the decisions that raters use while scoring compositions holistically, beginning trends to analyze raters' thinking and decision-making processes as validity evidence in L2 writing assessments, developed later in CUMMING ET AL. (2002) and LUMLEY (2005).	**E**

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

1988	Santos, T. (1988). Professors' reactions to the academic writing of non-native speaking students. *TESOL Quarterly* 22, 69–90.	In an effort to establish the most important criteria for evaluating L2 writing, **Santos** presented one among many studies that surveyed the attitudes of university professors when evaluating the writing of L2 students.	**B**, **D** & **E**
1988	Stansfield, C. & J. Ross (1988). A long-term research agenda for the Test of Written English. *Language Testing* 5, 160–186.	**Stansfield & Ross** described the rationale and a research agenda for the introduction of the Test of Written English as an assessment of writing performance supplementary to the TOEFL, responding to requests from educators wanting evidence of ESL students' communicative performance in writing, and the unfulfilled promise of CARROLL (1961) for a full design of the TOEFL.	**B**, **C**, **D** & **E**
1989	Hudelson, S. (1989). *Write on: Children writing in ESL.* Englewood Cliffs, NJ: Prentice Hall Regents.	Many previously neglected issues in the teaching and assessment of young children's writing in an L2 in school contexts were discussed by **Hudelson**, with concrete recommendations to link teaching, learning, and assessment while promoting bilingual abilities.	**A** & **C**
1988	Ruth, L. & S. Murphy (1988). *Designing writing tasks for the assessment of writing.* Norwood, NJ: Ablex.	**Ruth & Murphy** highlighted the importance of task variables in writing assessment while providing exemplary methods and guidelines for educators to prepare and evaluate various types of tasks to evaluate students' writing.	**C** & **E**
1990	Bachman, L. (1990). *Fundamental considerations in language testing.* Oxford: Oxford University Press. Messick, S. (1989). Validity. In R. Linn (ed.), *Educational measurement* (3rd edn.). New York: Macmillan, 13–103.	**Bachman**'s authoritative analysis built on CANALE & SWAIN'S (1980) and others' appeals for communicative models of language assessment, notably the direct assessment of writing performance, rather than indirect tests of knowledge about writing. The book also introduced to language testing research **Messick**'s unitary theory of test validity, focusing on construct validity, interpretations and uses of test scores, and the long-term consequences of using tests. Multiple sources of evidence and methods of validation are needed to justify writing assessments and evaluate if they achieve their intended purposes.	**B** & **E**

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
1990	Raimes, A. (1990). The TOEFL Test of written English: Causes for concern. *TESOL Quarterly* 24, 427–442.	**Raimes** voiced many educators' concerns over the negative influences on teaching and learning of timed, short-essay tests required for university admissions. Her concerns for the backwash effects of formal L2 writing tests on educational practices have persisted until the present day.	**C** & **E**
1991	Hamp-Lyons, L. (ed.) (1991). *Assessing second language writing in academic contexts*. Norwood, NJ: Ablex.	**Hamp-Lyons** edited the first full-length book with chapters by several scholars defining key issues in the assessment of second language writing, such as the design of assessment tasks, choice of scoring methods, and inquiry into and training of raters' decision making.	**B** & **E**
1992	Purves, A. (ed.) (1992). *The IEA study of written composition: Education and performance in fourteen countries*. Oxford: Pergamon.	**Purves** and colleagues' survey of writing achievement and education internationally demonstrated that, despite common characteristics in written compositions, standards for the assessment of writing vary by culture and educational systems.	**A**, **B**, **C**, **D** & **E**
1996	Grabe, W. & R. Kaplan (1996). *Theory and practice of writing: An applied linguistic perspective*. Harlow: Longman.	In addition to reviewing theoretical and research perspectives on L2 writing, **Grabe & Kaplan** described teaching and assessment practices appropriate to learners at different levels of proficiency, such as simple completion of forms at beginning levels and complex research papers at advanced levels of proficiency.	**A**, **B** & **C**
1996	Calfee, R. & P. Perfumo (eds.) (1996). *Writing portfolios in the classroom: Policy and practice, promise and peril*. Mahwah, NJ: Lawrence Erlbaum.	**Calfee & Perfumo**'s volume consolidated many of the pedagogical practices, assessment challenges, and policy issues associated with portfolios as a medium for gathering information for writing assessments in classroom or tests.	**A**, **B** & **E**
1996	Li, X. (1996). *'Good writing' in cross-cultural context*. Albany, NY: State University of New York Press.	**Li** demonstrated, through detailed case studies in the U.S. and China, variation in concepts about assessing the quality of writing within as well as between countries and among highly regarded teachers. These analyses exemplify vividly issues arising from PURVES (1992).	**A**, **B**, **C**, **D** & **E**

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

1996	Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning* 46, 327–369.	Reviewing past research and pedagogical principles, **Truscott** argued that teachers' corrections of grammar in students' writing are ineffective, counter-productive, and even harmful, and so should not be used in classroom assessment or teaching.	**B** & **C**
1998	Brindley, G. (1998). Outcomes-based assessment and reporting in language learning programs. *Language Testing* 15, 45–85.	**Brindley** synthesized a range of research, curriculum innovations, and assessments based on competencies or benchmarks that, following international trends in education, specified intended outcomes for educational programs aiming integrally to link teaching, learning, and assessment. Examples were cited from the Certificates in Spoken and Written English, which in Australia provided an assessment framework that also defined competencies taught in the national curriculum for adult migrant learners and were based on theories of systemic-functional linguistics.	**B**, **C** & **E**
1998	Wolfe-Quintero, K., S. Inagaki & H-Y. Kim (1998). *Second language development in writing: Measures of fluency, accuracy and complexity*. Honolulu, HI: University of Hawaii at Manoa.	**Wolfe-Quintero et al.** synthesized and evaluated previous research on L2 writing demonstrating, for research purposes, the centrality and reliability of analyses of learners' texts that measure fluency, accuracy, and complexity, relating these to Peter Skehan's cognitive model of L2 skills.	**B**, **D** & **E**
2001	Cumming, A. (2001). ESL/EFL writing instructors' practices for assessment: General or specific purposes. *Language Testing* 18, 207–224.	Reporting results of interviews about the practices for assessing writing of experienced instructors of ESL and EFL in six countries, **Cumming** established a basis of commonalities as well as variability in how L2 composition is taught and evaluated (e.g. emphasizing language and style, rhetoric, composing processes, self-expression, or socialization).	**A**, **C** & **E**
2002	Cumming, A., R. Kantor & D. Powers. (2002). Decision making while assessing ESL/EFL writing: A descriptive framework. *Modern Language Journal* 86, 67–96.	**Cumming et al.** produced an empirical model of the decisions that raters make while evaluating ESL compositions for proficiency tests, distinguishing between thinking processes for interpretations or judgements of language, content, and rhetorical organization.	**B** & **E**

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

YEAR	REFERENCES	ANNOTATIONS	THEME
2002	Weigle, S. (2002). *Assessing writing*. Cambridge: Cambridge University Press.	**Weigle** published an authoritative and comprehensive book on major aspects of assessing second and foreign languages. The book serves as the major, recent synthesis of research and theories about L2 writing assessment for purposes of teacher development and test design.	**B**, **C** & **E**
2002	Liu, J. & J. Hansen. (2002). *Peer response in second language writing classrooms*. Ann Arbor, MI: University of Michigan Press.	**Liu & Hansen** synthesized research on and pedagogical practices for peer response (i.e. classmates responding to each others' written drafts) as an approach to assessment with pedagogical, learning, community building value in L2 classrooms.	**A**, **C** & **E**
2003	Ferris, D. (2003). *Response to student writing: Implications for second language students*. Mahwah, NJ: Lawrence Erlbaum.	**Ferris** synthesized her and others' research on teachers' methods of responding to students' L2 writing and related approaches to formative assessment, arguing against TRUSCOTT (1996).	**B**, **C** & **E**
2003	Shermis, M. & J. Burstein (eds.) (2003). *Automated essay scoring: A cross-disciplinary perspective*. Mahwah, NJ: Lawrence Erlbaum.	**Shermis & Burstein** documented recent advances and future challenges in the development of automated scoring techniques for the assessment of writing in L1 and L2s. They acknowledge present limitations in these methods, particularly given the variability of L2 writing, but show how methods of automated scoring are rapidly developing to produce writing assessments that are consistent (as human scoring) and follow explicit, though narrowly linguistic, criteria.	**A**, **B** & **E**
2003	Jarvis, S., L. Grant, D. Bikowski & D. Ferris (2003). Exploring multiple profiles of highly rated learner compositions. *Journal of Second Language Writing* 12, 377–403.	**Jarvis et al.** demonstrated that the compositions written by L2 students with high scores on a test tend to display differing, rather than common, clusters of language features. Holistic scoring of writing is therefore flexible in encompassing differing trajectories in students' writing development.	**B**, **D** & **E**
2005	Schoonen, R. (2005). Generalizability of writing scores: An application of structural equation modeling. *Language Testing* 22, 1–30.	**Schoonen** applied generalizability theory to determine the adequacy of samples of writing and of raters for assessment purposes. As with other studies of this kind, more samples of students' writing increased the validity of assessments compared to using more raters to score compositions.	**B** & **E**

https://doi.org/10.1017/S0261444808005430 Published online by Cambridge University Press

2005	Lumley, T. (2005). *Assessing second language writing: The rater's perspective*. Frankfurt am Main: Peter Lang.	**Lumley** reviewed research, including his own, on raters' decision making while evaluating ESL compositions. He argued that individual raters' knowledge and the policies and purposes of tests are more central to raters' actual uses of scoring scales than are the descriptive criteria in the scales.	**B**, **C** & **E**
2005	Alderson, J. C. (2005). *Diagnosing foreign language proficiency*. London: Continuum.	**Alderson** argued for the value of assessing language abilities for diagnostic and self-assessment purposes, documenting studies for the DIALANG Internet-based scheme in fourteen European languages, including analytic rating of writing abilities in the form of 'can do' statements.	**A**, **B**, **C** & **E**
2005	Goldstein, L. (2005). *Teacher written commentary in second language writing classrooms*. Ann Arbor, MI: University of Michigan Press.	**Goldstein** reviewed research and pedagogical practices for responding to the writing of L2 students, taking a more holistic and ethnographic approach than appears in FERRIS (2003) but still arguing against the position advocated by TRUSCOTT (1996). Goldstein recommends judicious and informed feedback on content and language in classroom writing assessments.	**A**, **B** & **C**
2006	Hyland, K. & F. Hyland (eds.) (2006). *Feedback in second language writing: Contexts and issues*. New York: Cambridge University Press.	**Hyland & Hyland** presented various perspectives on the assessment and provision of feedback on students' L2 writing, expanding on ideas in FERRIS (2003) and GOLDSTEIN (2005).	**A**, **B**, **C** & **D**
2007	Barkaoui, K. (2007). Participants, texts, and processes in ESL/EFL essay tests: A narrative review of the literature. In A. Cumming & M. Laurier (eds.), *Language assessment*. Special issue of *The Canadian Modern Language Review* 64.1, 99–134.	Synthesizing previous empirical studies of ESL/EFL writing tests, **Barkaoui** demonstrated how they have focused on the effects of tasks, essays, or raters rather than on much systematic consideration of examinee characteristics, scoring methods, or assessment contexts.	**B** & **E**
2008	Chapelle, C., M. Enright & J. Jamieson (eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language*. New York: Routledge.	**Chapelle et al.** documented the development, research, and validation processes in revisions to the TOEFL to produce the new TOEFL-iBT, including independent and integrated writing tasks.	**B**, **C** & **E**