Lang. Teach. (2013), **46.1**, 41–52 © Cambridge University Press 2012
doi:10.1017/S0261444812000377

# Research Timeline

## Second language vocabulary assessment

**John Read** University of Auckland, New Zealand
ja.read@auckland.ac.nz

The focus of this timeline article is on the development and validation of measures to assess the vocabulary knowledge and ability of second language (L2) learners. There is a parallel and somewhat longer tradition of work on L1 vocabulary assessment, particularly by reading researchers in the United States. Some key references from the L1 research prior to the 1980s have been included in the present timeline, mainly on the basis that they had a strong influence on the thinking of L2 researchers as they began to work on their own measures of vocabulary knowledge. Since that time, L1 and L2 research have continued along largely parallel tracks, with regrettably little interaction between the two (for an account of the state of the art in L1 vocabulary assessment, see Pearson et al. 2007).

Obviously the inclusion in examinations of questions and test items to assess learners' knowledge of foreign language vocabulary goes back much further than the 1980s. But the argument here is that it was during that decade that researchers began to publish reports of particular measures and procedures designed to assess defined aspects of L2 vocabulary knowledge or ability, rather than simply to test whether learners knew sets of words that they were supposed to have studied.

The decade of the 1980s is significant in that during that period (as has been frequently observed) a small band of dedicated researchers was working in comparative isolation from the mainstream of two disciplines in which one might have expected them to have a more central role: second language acquisition (SLA) research and language testing. SLA was then, and is still, predominantly concerned with the acquisition by learners of syntactic and morphological features of the target language. At the same time, language testing was under the influence of the communicative approach to language teaching, and the focus of attention in the literature shifted from knowledge of the lexical and grammatical elements of the language to the performance of language use tasks. Thus, in both cases, work on vocabulary was not seen as a priority.

At the conceptual level, the result of this isolation from wider influences was a relative lack of theory to account for the nature of L2 vocabulary knowledge and the processes of lexical development in L2. From a testing perspective, it meant that there was only a limited basis for defining the constructs being measured by particular language tests. Distinctions such as receptive vs. productive knowledge, or breadth and depth of knowledge, were made as if they were unproblematic and could be clearly understood without much further elaboration. Paul Meara has stood out among scholars working on L2 vocabulary development as one who has long pursued a more theoretical approach informed by psycholinguistics,

https://doi.org/10.1017/S0261444812000377 Published online by Cambridge University Press

employing computer simulations and modelling along with a range of innovative measures to operationalise the theoretical concepts (Meara 2002, 2009).

Apart from the work of Meara and his associates, the timeline shows how a number of authors in the last fifteen years have sought to identify various components or dimensions of vocabulary knowledge and to apply different approaches to the definition of the construct.

At the operational level, the separation from language testing meant that early papers in the 1980s and 1990s presented rather limited evidence for the quality of the test instruments. They relied typically on classical reliability statistics and correlations with concurrent measures, as well as *t*-tests or simple analyses of variance where appropriate. There was quite a lag before the principles of modern validation theory – which had been widely disseminated in the field of language testing through Bachman's (1990) landmark volume – were reflected in research articles on L2 vocabulary tests. In the last ten years more sophisticated procedures, notably Rasch analysis, have been applied to the evaluation of vocabulary instruments, along with more complex and well-formulated validity arguments in some recent cases. In addition, there have been some critical re-evaluations of widely known vocabulary measures, to take a fresh look at their conceptual and empirical adequacy. It is symptomatic of the return of vocabulary assessment to the mainstream that there has been a noticeable increase in the number of articles on the topic accepted for publication in the major research journal *Language Testing*.

The selection of publications here is based on the author's extensive knowledge of the field, cross-checked against books found in the timeline itself – Read (2000), Milton (2009), Schmitt (2010), Nation & Webb (2011) – and a search of recent volumes of the major journals in applied linguistics.

The items in the timeline represent the following themes, which are labelled with the appropriate letter(s) in the last column of the table:

**A** Vocabulary size (breadth)
**B** Depth of vocabulary knowledge
**C** Knowledge of spoken vocabulary
**D** Productive use of vocabulary in speech and writing
**E** Role of context in vocabulary assessment
**F** Speed of lexical access
**G** Definition of the vocabulary construct
**H** Further validation or critical analysis of an existing vocabulary measure

## References

Bachman, L. F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
Meara, P. (2002). The rediscovery of vocabulary. *Second Language Research* 18.4, 393–407.
Meara, P. (2009). *Connected words: Word associations and second language vocabulary acquisition*. Amsterdam: John Benjamins.
Pearson, P. D., E. H. Hiebert & M. L. Kamil (2007). Vocabulary assessment: What we know and what we need to know. *Reading Research Quarterly* 42.2, 282–296.

JOHN READ is an Associate Professor of Applied Language Studies at the University of Auckland, where he conducts research and teaches courses in language assessment, and vocabulary teaching and learning. He is the author of *Assessing vocabulary* (Cambridge, 2000), was co-editor of *Language Testing* (2002–2006), and has been President of the International Language Testing Association (2011–2012).

https://doi.org/10.1017/S0261444812000377 Published online by Cambridge University Press

https://doi.org/10.1017/S0261444812000377 Published online by Cambridge University Press



| YEAR | REFERENCES | ANNOTATIONS | THEME/CONCEPT |
|---|---|---|---|
| 1935 | Stalnaker, J. M. & W. Kurath (1935). A comparison of two types of foreign language vocabulary test. *Journal of Educational Psychology* 26, 435–442. | **Stalnaker & Kurath** assessed American university students' knowledge of 100 German words in isolation (matching the word with its English equivalent) and in the context of a specially written reading passage in German (supplying the English equivalent of each target word). Since the results were very similar, they concluded that both formats were valid measures of vocabulary knowledge. | **E** |
| 1942 | Cronbach, L. J. (1942). An analysis of techniques for diagnostic vocabulary testing. *Journal of Educational Research* 36.3, 206–217. | In what can be seen as an early analysis of vocabulary depth (although he did not use the term), **Cronbach** identified five types of word knowledge – generalisation, application, breadth of meaning, precision of meaning and availability – and discussed the test formats that could be appropriately be used to assess each type. | **B, G** |
| 1963 | Lorge, I. & J. Chall (1963). Estimating the size of vocabularies of children and adults: An analysis of methodological issues. *Journal of Experimental Education* 32.2, 147–157. | **Lorge & Chall** discussed a number of methodological issues that helped explain the widely differing estimates of vocabulary size obtained in previous studies. The issues were associated with the procedure for obtaining a vocabulary sample, the definition of a word, and the method of testing word knowledge. | **A** |
| 1981 | Anderson, R. C. & P. Freebody (1981). Vocabulary knowledge. In J. T. Guthrie (ed.), *Comprehension and teaching: Research reviews*. Newark, DE: International Reading Association, 77–117. | In a wide-ranging discussion of the role of vocabulary knowledge in reading comprehension, **Anderson & Freebody** introduced the distinction between breadth and depth of knowledge. They went on to discuss issues in obtaining a valid measure of vocabulary size (or breadth). | **A, B** |
| 1983 | Anderson, R. C. & P. Freebody (1983). Reading comprehension and the acquisition of word knowledge. In B. Hutson (ed.), *Advances in reading/language research* (vol. 2). Greenwich, CT: JAI Press, 231–256. | As part of their L1 reading research, **Anderson & Freebody** were the first scholars to use for educational purposes a Yes/No test that incorporated non-words and applied a scoring formula to adjust for over-reporting of knowledge. They obtained evidence that it was a more valid measure of word knowledge than a multiple-choice test. | **A** |

https://doi.org/10.1017/S0261444812000377 Published online by Cambridge University Press



| YEAR | REFERENCES | ANNOTATIONS | THEME/CONCEPT |
|---|---|---|---|
| 1983 | Nation, P. (1983). Testing and teaching vocabulary. *Guidelines* 5.1, 12–25. | In a language teachers' journal published in Singapore, **Nation** presented the original version of the Vocabulary Levels Test, which was to become arguably the most influential L2 English vocabulary test in the world. It was offered as a useful diagnostic tool for teachers in the classroom. | **A** |
| 1987 | Meara, P. & B. Buxton (1987). An alternative to multiple choice vocabulary tests. *Language Testing* 4.2, 142–154. | **Meara & Buxton** were the first authors to publish research on an English vocabulary test for L2 learners that employed the Yes/No format as a measure of vocabulary size, drawing on the work of ANDERSON & FREEBODY[1] (1983). | **A** |
| 1988 | Read, J. (1988). Measuring the vocabulary knowledge of second language learners. *RELC Journal* 19.2, 12–25. | In an early validation study of NATION'S (1983) test, **Read** showed how scores changed from the beginning to the end of a three-month intensive academic English course. He also investigated the implicational scaling of scores at the five vocabulary levels represented in the test. | **A, H** |
| 1990 | Goulden, R., P. Nation & J. Read (1990). How large can a receptive vocabulary be? *Applied Linguistics* 11.4, 341–363. | **Goulden et al.** drew on earlier work by L1 researchers (e.g. LORGE & CHALL 1963) to obtain a valid sample of words from an unabridged dictionary to use as the basis for five forms of a self-assessment task to measure vocabulary size. | **A** |
| 1991 | Henning, G. (1991). *A study of the effects of contextualization and familiarization on responses to the TOEFL vocabulary items*. TOEFL Research Reports, 35. Princeton, NJ: Educational Testing Service. | In response to criticism that the vocabulary items in the Test of English as a Foreign Language (TOEFL) were too discrete in nature, **Henning** investigated various more contextualised multiple-choice formats and found that the best one overall presented the target words embedded in a reading passage, rather than in individual sentences. | **E** |
| 1993 | Read, J. (1993). The development of a new measure of L2 vocabulary knowledge. *Language Testing* 10.3, 355–371. | Building on earlier work by Paul Meara on the standard word association task, **Read** devised the Word Associates Format as a way to measure vocabulary depth. He developed a test to assess knowledge of academic vocabulary through associative links between a target word and other related words. | **B** |

https://doi.org/10.1017/S0261444812000377 Published online by Cambridge University Press



| YEAR | REFERENCES | ANNOTATIONS | THEME/CONCEPT |
|---|---|---|---|
| 1993 | Verhallen, M. & R. Schoonen (1993). Lexical knowledge of monolingual and bilingual children. *Applied Linguistics* 14.4, 344–363. | In their research on the depth of vocabulary knowledge of primary school children in the Netherlands, **Verhallen & Schoonen** employed a structured interview procedure to elicit all the 'meaning aspects' of six high-frequency Dutch nouns. Bilingual immigrant children produced significantly fewer meaning aspects than did monolingual Dutch children. | **B** |
| 1995 | Laufer, B. & P. Nation (1995). Vocabulary size and use: Lexical richness in L2 written production. *Applied Linguistics* 16.3, 307–322. | In order to measure productive vocabulary knowledge, in the sense of learners' ability to use words in their own speech and writing, **Laufer & Nation** created the Lexical Frequency Profile to calculate the relative percentages of high-frequency, academic and lower-frequency vocabulary in learners' written compositions. | **D** |
| 1996 | Wesche, M. B. & T. S. Paribakht (1996). Assessing second language vocabulary knowledge: Depth vs. breadth. *Canadian Modern Language Review* 53.1, 13–40. | **Wesche & Paribakht** developed as a measure of depth of knowledge their Vocabulary Knowledge Scale, which elicits learner ratings of how well they know the meaning of a set of target words and, where applicable, requires them to supply verifying evidence in the form of a synonym, translation or the use of the word in a sentence. | **B** |
| 1998 | Read, J. (1998). Validating a test to measure depth of vocabulary knowledge. In A. J. Kunnan (ed.), *Validation in language assessment*. Mahwah, NJ: Laurence Erlbaum, 41–60. | Building on his earlier work (READ 1993), **Read** devised a version of the Word Associates Format that was designed specifically to assess depth of knowledge of high-frequency adjectives in English. Evidence for the validity of the test was obtained by using two concurrent measures: an interview task and a specially created multiple-option gap-filling test. | **B** |
| 1999 | Beglar, D. & A. Hunt (1999). Revising and validating the 2000 Word Level and University Word Level Vocabulary Tests. *Language Testing* 16.2, 131–162. | In a study that foreshadowed the one by SCHMITT ET AL. (2001), **Beglar & Hunt** produced revised forms of the Vocabulary Levels Test (NATION 1983) at just the 2000-word and University word levels, and validated the reduced test as a measure of the English vocabulary knowledge of Japanese learners. | **A, H** |

https://doi.org/10.1017/S0261444812000377 Published online by Cambridge University Press



| YEAR | REFERENCES | ANNOTATIONS | THEME/CONCEPT |
|---|---|---|---|
| 1999 | Henriksen, B. (1999). Three dimensions of vocabulary development. *Studies in Second Language Acquisition* 21.2, 303–317. | In contrast to the static or cross-sectional accounts that have predominated in work on vocabulary assessment, **Henriksen** proposed that the development of vocabulary knowledge proceeds along three dimensions, which she identified as partial to precise knowledge, depth of knowledge (developing a network of links between words), and receptive to productive knowledge. | **B, G** |
| 1999 | Laufer, B. & P. Nation (1999). A vocabulary-size test of controlled productive ability. *Language Testing* 16.1, 33–51. | Taking a different approach from their Lexical Frequency Profile (LAUFER & NATION 1995), **Laufer & Nation** measured 'productive' vocabulary by means of sentence-based gap-filling items. The test used the same target words as the original Vocabulary Levels Test (NATION 1983), so that the relative size of receptive and productive vocabulary knowledge could be calculated. | **A** |
| 1999 | Qian, D. D. (1999). Assessing the roles of depth and breadth of knowledge in reading comprehension. *Canadian Modern Language Review* 56.2, 282–308. | Using NATION's (1983) Vocabulary Levels Test and READ's (1998) Word Associates Test for breadth and depth respectively, **Qian** found that the depth test could account for a significant amount of the variance in reading comprehension scores, in addition to what was measured by the breadth test. | **A, B** |
| 2000 | Bogaards, P. (2000). Testing L2 vocabulary knowledge at a high level: The case of the Euralex French tests. *Applied Linguistics* 21.4, 490–516. | **Bogaards** reviewed work to date on the measurement of depth and then reported on a study to investigate the validity of Paul Meara's Euralex French Tests – a simplified version of READ's (1993) Word Associates Format – for advanced learners of the language. | **B** |
| 2000 | Fountain, R. L. & I. S. P. Nation (2000). A vocabulary-based graded dictation test. *RELC Journal* 31.2, 29–44. | This article by **Fountain & Nation** is highlighted as one of very few that reports on the assessment of spoken vocabulary, for placement purposes. The specially written texts are graded in terms of progressively lower frequency words in each succeeding paragraph, and only the target content words are scored. | **C** |

https://doi.org/10.1017/S0261444812000377 Published online by Cambridge University Press



| YEAR | REFERENCES | ANNOTATIONS | THEME/CONCEPT |
|---|---|---|---|
| 2000 | Meara, P. & T. Fitzpatrick (2000). An improved method of assessing productive vocabulary in L2. *System* 28.1, 19–30. | **Meara & Fitzpatrick** devised Lex30 as an efficient measure of productive vocabulary knowledge based on word associations supplied by learners in response to 30 stimulus words. It produced promising preliminary results. | **D** |
| 2000 | Read, J. (2000). *Assessing vocabulary*. Cambridge: Cambridge University Press. | **Read**'s volume was the first book on second language vocabulary assessment. It sets out to give a comprehensive account of vocabulary testing from the perspective of modern language testing theory. | **A, B, C, D, E, G** |
| 2001 | Beeckmans, R., J. Eyckmans, V. Janssens, M. Dufranne & H. van de Velde (2001). Examining the Yes/No vocabulary test: Some methodological issues in theory and practice. *Language Testing* 18.3, 235–274. | Based on their work with French-speaking learners of Dutch in Belgium, **Beeckmans et al.** took a critical perspective on the Yes/No test, raising questions about how to take account of response bias in scoring the test and how to calculate its reliability. | **A, H** |
| 2001 | Greidanus, T. & L. Nienhuis (2001). Testing the quality of word knowledge in a second language by means of word associations: Types of distractors and types of associations. *The Modern Language Journal* 85.4, 567–577. | **Greidanus & Nienhuis** applied READ's (1993) Word Associates Format to assessing learners of a language other than English: advanced students of French in the Netherlands. They explored the use of semantically related distractors and the relative difficulty of different types of associates. | **D, H** |
| 2001 | Laufer, B. & P. Nation (2001). Passive vocabulary size and speed of meaning recognition. *EUROSLA Yearbook* 1, 7–28. | This appears to be the first study to measure the *speed* as well as the accuracy of L2 learners' responses to vocabulary test items. **Laufer & Nation** created a computer-based version of the Vocabulary Levels Test (NATION 1983) for the purpose. | **A, F** |

https://doi.org/10.1017/S0261444812000377 Published online by Cambridge University Press



| YEAR | REFERENCES | ANNOTATIONS | THEME/CONCEPT |
|---|---|---|---|
| 2001 | Meara, P. & H. Bell (2001). P_Lex: A simple and effective way of describing the lexical characteristics of short texts. *Prospect* 16.3, 5–19. | **Meara & Bell** created P_Lex as a measure of learners' productive vocabulary use that would be more reliable than LAUFER & NATION's (1995) Lexical Frequency Profile, especially with low-proficiency learners who produced only a few non-high-frequency words. | **D** |
| 2001 | Nation, I. S. P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press. | Although **Nation**'s comprehensive volume is not primarily concerned with assessment, it includes an influential framework of what is involved in knowing a word and contains a chapter that discusses how each aspect of word knowledge can be assessed. | **A, B, C, D, E, G** |
| 2001 | Read, J. & C. A. Chapelle (2001). A framework for second language vocabulary assessment. *Language Testing* 18.1, 1–32. | **Read & Chapelle** applied modern validation theory to the analysis of vocabulary assessments, showing in particular how an analysis of the purpose of a test could be systematically related to its design features. They drew on Chapelle's work on construct definition as well as READ's (2000) framework for classifying vocabulary measures. | **G** |
| 2001 | Schmitt, N., D. Schmitt & C. Clapham (2001). Developing and exploring the behaviour of new versions of the Vocabulary Levels Test. *Language Testing* 18.1, 55–88. | **Schmitt et al.** undertook a comprehensive validation study of a substantially revised version of the Vocabulary Levels Test (NATION 1983), involving over 800 learners in five different countries, as well as multiple forms of validity evidence. | **A, H** |
| 2002 | Cameron, L. (2002). Measuring vocabulary size in English as an additional language. *Language Teaching Research* 6.2, 145–173. | **Cameron**'s study stands out as involving secondary school students with English as an additional language in the UK rather than older learners. NATION's (1983) Vocabulary Levels Test proved to be a much more satisfactory measure than a Yes/No test, and it showed a more uneven profile of knowledge across the frequency levels than that found in READ (1988). | **A, H** |

https://doi.org/10.1017/S0261444812000377 Published online by Cambridge University Press



| YEAR | REFERENCES | ANNOTATIONS | THEME/CONCEPT |
|---|---|---|---|
| 2002 | Huibregtse, I., W. Admiraal & P. Meara (2002). Scores on a yes-no vocabulary test: Correction for guessing and response style. *Language Testing* 19.3, 227–245. | Complementing the work of BEECKMANS ET AL. (2001), **Huibregtse et al.** discussed four different scoring formulae for the Yes/No test, from the perspective of both correcting for guessing and taking account of test-taker response styles. | **A** |
| 2004 | Laufer, B., C. Elder, K. Hill & P. Congdon (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing* 21.2, 202–226. | Using items from SCHMITT ET AL.'s (2001) Vocabulary Levels Test, **Laufer et al.** designed tasks to assess different 'strengths' of knowledge of the form–meaning link, through active and passive recognition and recall of the words. Rasch analysis was used to explore the relationship between the measures of size and strength of knowledge. | **A** |
| 2004 | Malvern, D. D., B. J. Richards, N. Chipere & P. Durán (2004). *Lexical diversity and language development: Quantification and assessment*. Basingstoke: Palgrave Macmillan. | **Malvern et al.** present a comprehensive account of their work to develop and validate D as a measure of lexical diversity in the spoken and written production of L1 and L2 language users. It was designed as a more valid substitute for established measures such as the Type–Token Ratio. | **D** |
| 2004 | Read, J. (2004). Plumbing the depths: How should the construct of vocabulary knowledge be defined? In P. Bogaards & B. Laufer (eds.), *Vocabulary in a second language: Selection, acquisition, and testing*. Amsterdam: John Benjamins, 209–227. | Covering some of the same ground as HENRIKSEN (1999), **Read** reviewed work to date on conceptualising and measuring depth of vocabulary knowledge in terms of three approaches: precision of meaning, comprehensive word knowledge and network knowledge. He argued that the single term 'depth' was inadequate to represent all the aspects of word knowledge involved. | **B, G** |

https://doi.org/10.1017/S0261444812000377 Published online by Cambridge University Press



| YEAR | REFERENCES | ANNOTATIONS | THEME/CONCEPT |
|---|---|---|---|
| 2004 | Snellings, P., A. van Gelderen & K. de Klopper (2004). Validating a test of second language written lexical retrieval: A new measure of fluency in written language production. *Language Testing* 21.2, 174–201. | **Snellings et al.** devised the computer-based Written Productive Translation Test (WPPT) to measure the learners' speed of retrieval of L2 words. The learners have to type the English equivalents of Dutch words, which are presented in a short English sentence. | **F** |
| 2005 | Meara, P. (2005). Lexical frequency profiles: A Monte Carlo analysis. *Applied Linguistics* 26.1, 32–47.<br>Laufer, B. (2005). Lexical frequency profiles: From Monte Carlo to the real world. *Applied Linguistics* 26.4, 582–588. | **Meara** published a critical analysis of LAUFER & NATION's (1995) Lexical Frequency Profile, using multiple computer simulations to raise questions about its reliability and validity as a measure of productive vocabulary knowledge of learners at different proficiency levels. **Laufer** responded with a vigorous defence of research based on real data from language learners, and argued that Meara had misrepresented the original study in various ways. | **D, H** |
| 2005 | Zareva, A., P. Schwanenflugel & Y. Nikolova (2005). Relationship between lexical competence and language proficiency: Variable sensitivity. *Studies in Second Language Acquisition* 27.4, 567–595. | **Zareva et al.** added metacognitive awareness as a third dimension, along with breadth and depth, in their cross-sectional investigation of the lexical competence of L1 and L2 users of English, using a modified version of WESCHE & PARIBAKHT's (1996) Vocabulary Knowledge Scale, plus a word association task to measure metacognitive awareness. | **A, B** |
| 2006 | Milton, J. & N. Hopkins (2006). Comparing phonological and orthographic vocabulary size: Do vocabulary tests underestimate the knowledge of some learners? *Canadian Modern Language Review* 63.1, 127–147. | Like FOUNTAIN & NATION (2000), this is one of the handful of studies to investigate spoken vocabulary. **Milton & Hopkins** developed an aural version of the Yes/No test and compared the performance of Greek- and Arabic-speaking learners on the aural and written versions. | **A, C** |

https://doi.org/10.1017/S0261444812000377 Published online by Cambridge University Press



| YEAR | REFERENCES | ANNOTATIONS | THEME/CONCEPT |
|---|---|---|---|
| 2007 | Daller, H., J. Milton & J. Treffers-Daller (eds.) (2007). *Modelling and assessing vocabulary knowledge*. Cambridge: Cambridge University Press. | The volume edited by **Daller et al.** brings together important work by European scholars on a variety of topics in vocabulary assessment, but particularly on the measurement of the lexical richness of learners' spoken and written production. | **A, B, D, H** |
| 2009 | Barfield, A. & H. Gyllstad (eds.) (2009). *Researching collocations in another language: Multiple interpretations*. New York: Palgrave Macmillan. | One section of this volume edited by **Barfield & Gyllstad** contains three chapters by Revier, Eyckmans & Gyllstad, in which the authors present test formats which they have devised to assess learners' knowledge of verb-noun collocations. Very little other work has been done on how to assess knowledge of collocations or other multi-word lexical units. | **B** |
| 2009 | Bruton, A. (2009). The Vocabulary Knowledge Scale: A critical analysis. *Language Assessment Quarterly* 6.4, 288–297. | **Bruton** conducted a systematic analysis of the Vocabulary Knowledge Scale (WESCHE & PARIBAKHT 1996) and found that it had numerous shortcomings, in both conceptual and measurement terms. | **B, H** |
| 2009 | Harrington, M. & M. Carey (2009). The online Yes/No test as a placement tool. *System* 37.4, 614–626. | **Harrington & Carey** looked at speed as well as accuracy of response of learners to a Yes/No test used as part of a placement battery in a language school. The response time data was consistent with the accuracy scores, but less sensitive to differences in proficiency level. | **A, F** |
| 2009 | Milton, J. (2009). *Measuring second language vocabulary acquisition*. Bristol, UK: Multilingual Matters. | This second (after READ 2000) book-length treatment of vocabulary assessment by **Milton** gives particular attention to growth in vocabulary size by learners under various contextual influences, but also covers depth of knowledge and measures of lexical richness in production. | **A, B, C, D** |
| 2010 | Beglar, D. (2010). A Rasch-based validation of the Vocabulary Size Test. *Language Testing* 27.1, 101–118. | **Beglar** undertook the first large-scale validation of Nation & Beglar's Vocabulary Size Test with learners of English in Japan, using sophisticated Rasch techniques. The test, which uses a multiple-choice format, samples words from the first 14,000 word families in the British National Corpus. An introductory article on the test by Nation & Beglar (2007) can be found in *The Language Teacher* 31.7, 9–13. | **A, H** |

https://doi.org/10.1017/S0261444812000377 Published online by Cambridge University Press



| YEAR | REFERENCES | ANNOTATIONS | THEME/CONCEPT |
|---|---|---|---|
| 2010 | Fitzpatrick, T. & J. Clenton (2010). The challenge of validation: Assessing the performance of a test of productive vocabulary. *Language Testing* 27.4, 537–554. | Reviewing a number of research studies conducted on the Lex30 test (MEARA & FITZPATRICK 2000), **Fitzpatrick & Clenton** make a case for the reliability and construct validity of the test, while recognising its limitations. | **D, H** |
| 2010–11 | Schmitt, N. (2010). *Researching vocabulary: A vocabulary research manual*. Basingstoke, UK: Palgrave Macmillan. Nation, I. S. P. & S. Webb (2011). *Researching and analyzing vocabulary*. Boston, MA: Heinle. | These two books, by **Schmitt** and **Nation & Webb**, are grouped together, partly because they appeared almost at the same time but, more importantly, they have a similar aim, that of providing a comprehensive guide to novice researchers on how to conduct research on vocabulary. The design and use of appropriate measures of vocabulary knowledge and ability is an integral part of the discussion in both books. | **A, B, C, D, E, F, H** |
| 2011 | Crossley, S. A., T. Salsbury, D. S. McNamara & S. Jarvis (2011). Predicting lexical proficiency in language learner texts using computational indices. *Language Testing* 28.4, 561–580. | Numerous recent studies by **Crossley et al.** have made use of Coh-Metrix, a computational tool which provides multiple linguistic and psycholinguistic indices of the lexical features of spoken and written texts. This study investigated the extent to which the indices could predict human ratings of the quality of written learner texts. | **D** |
| 2011 | Schmitt, N., J. W. C. Ng & J. Garras (2011). The word associates format: Validation evidence. *Language Testing* 28.1, 105–126. | **Schmitt et al.** conducted two studies to investigate the validity of READ's (1998) version of the Word Associates Format by means of a concurrent interview procedure and think-aloud protocols. They also examined the effects of varying the scoring method, the number of distractors and the type of distractors. | **B, H** |
| 2012 | Read, J. (2012). Piloting vocabulary tests. In G. Fulcher & F. Davidson (eds.). *The Routledge handbook of language testing*. London: Routledge, 307–320. | **Read** undertook a comprehensive review of the research on the Word Associates Format as a case study to illustrate what is involved in piloting a vocabulary test before it is used for assessment or research purposes. | **B** |

[1] Authors' names are shown in small capitals when the study referred to appears in this timeline.