Title of Project:
Exploring the Validity Evidence of the TOEFL IBT
Reading Test from a Cognitive Perspective

Researcher:
Hyo Jung Lim
Michigan State University
hyojung@msu.edu

Research Supervisor:
Dr. Paula Winke

Hyo Jung Lim

Project Summary:

The current study explores the validity of the TOEFL iBT reading test by examining test takers' reading processes. The study consists of three sub-studies led by the following research questions:

1) To what extent does the TOEFL iBT reading test measure the reading construct?
2) To what extent does the TOEFL iBT reading test activate the cognitive processes being measured?

3) To what extent does test format affect students' test scores and test-taking processes?

The eye-tracking study, which was conducted to answer the second and third research questions, aimed to reveal whether the high-stakes, standardized second/foreign language reading test elicits the same type and level of cognitive processing from test takers as real-world academic reading tasks and whether test takers actually deploy required reading skills as assumed. I believe this investigation could be a meaningful, empirical attempt in that the study takes test takers' minds under consideration for the purpose of test validation. To date, test validation in language testing has heavily relied on correlation-based statistical analysis, which only concerns test outcomes. This statistical approach is limited because numbers do not carry conceptual information, and test scores do not inform us about how test takers derive their answers. That being said, the cognitive processing approach can complement the conventional, statistic-based test validation methods while illuminating the test takers' processes during a test.

Ninety Chinese ESL students attending a large Midwestern university participated in the experiment. The participants took two reading tests, one with multiple-choice questions and the other with open-ended questions, on a computer whose screen was fitted with a Tobii TX300

eye-tracker (www.tobii.com). Their eye movements were recorded while taking the test. Participants also completed a vocabulary test, a grammar test, a lexical processing task, a sentence processing task, a working memory task, a strategy questionnaire, and a stimulated recall interview. The collected data were analyzed in three different ways to answer the above research questions.

The first analysis aimed to investigate the knowledge source or skills of learners that the TOEFL iBT reading test taps into. The reading comprehension scores were regressed on vocabulary knowledge, grammar knowledge, word processing skill, sentential processing skill, working memory, reading strategies, test management strategies, and test-wise strategies. The results showed that vocabulary and grammar knowledge and word processing skills were the most important predictors. Learners' use of test-wise strategies was, however, negatively correlated with test scores; that is, those who used test-taking tricks were likely to score low.

The second analysis aimed to examine learners' reading processes by tracking their eye movements under testing conditions. First, the predetermined criteria of Bax and Weir (2012) were used to analyze high scorers' eye recordings (e.g., gaze plots). Then, successful readers' eye movements (e.g., total fixation duration) were compared to those of unsuccessful readers item by item, as in Bax (2013). To identify reading types, reading time and angles between fixation points were considered. The following patterns emerged: expeditious reading skills (i.e., those skills used to read text quickly) were rarely activated, especially when skimming across paragraphs was anticipated (e.g., learning to read questions). Instead, careful reading (e.g., reading to learn or reading to memorize) at a sentential level seemed to prevail. For vocabulary questions that aimed to measure learners' abilities to make inferences, test takers seemed to rely on their stored knowledge. In regard to the comparisons of successful and unsuccessful readers, successful readers were not always faster at reading and locating key information. In cases in which group differences appeared in the eye-movement comparisons, the item types were often limited to vocabulary and factual questions, which primarily demand locating and matching skills at the lexical level.

The third analysis was aimed at inspecting the test format effect on test-taking strategies and test scores. Given that learners' use of test-taking strategies is induced by test format (Rogers & Harley, 1999), the presence of a test format effect was assumed to evidence the impact of test-taking strategies on test results. First, the reading test scores were regressed on the test format, the text effect, and the group difference factors. The results showed that students were likely to score higher on the multiple-choice questions than on the open-ended questions. Then, the eye movements of the participants who received a full score on the open-ended questions were compared to those of the participants who correctly answered the corresponding multiple-choice questions. The eye-tracking analysis showed that it took more time for test takers to read a text, a

question stem, and the key phrases in the text when open-ended questions were given. Additionally, the presence (or absence) of response choices seemed to alter the focus of test takers' attention on the text. Strategic reading behaviors were more evident for the multiple-choice questions than the open-ended questions.

The results of the three different analyses did not consistently support the validity of the TOEFL iBT reading test. Given that learners who had a large vocabulary, advanced knowledge of grammar, and rapid recognition likely scored higher than those who did not, the test seems to properly tap into learners' construct-relevant cognitive resources. However, careful scrutiny of test takers' reading processes revealed that, in general, the test items failed to elicit learners' expeditious reading skills, which contradicts the original intention of the test. Second language reading researchers have voiced their concerns about the importance of L2 reading fluency in the post-secondary academic setting (e.g., Grabe, 2010). That being said, the current study provides some issues for language testers to ponder upon, including how to realize the assessment of learners' fast reading skills in high-stakes, standardized tests especially at the global level and the extent to which test tasks can (or should) mimic real-world academic reading tasks. With regard to the effect of test takers' strategies on test performance, the first analysis suggested that there was little or no effect. However, the test format effects found in the third analysis alluded to the influence of learners' test-taking strategies, specific to the multiple-choice questions, on their test-taking processes and test scores.

The study neither confirms nor discredits the validity of the TOEFL iBT reading test. Rather, it emphasizes the need to more closely examine learners' test-taking processes for test validation rather than only examining the outcomes. As the eye movement data show, test takers do not necessarily take the intended path to capitalize on the intended cognitive resources. Bachman (1990) argued that all types of validity complement one another, each providing different pieces of evidence to justify test scores or uses. Thus, various types of empirical (both qualitative and quantitative) methods should be encouraged for test validation.

References

Abbott, M. L. (2006). ESL Reading strategies: Differences in arabic and mandarin speaker test performance. *Language Learning*, *56*(4), 633–670.

Alderson, J. C. (1984). Reading in a foreign language: A reading problem or a language problem? In J. C. Alderson & A. H. Urquhart (Eds.), *Reading in a foreign language* (pp. 1–27). London,UK: Longman.

Alderson, J.C. (2000). *Assessing reading*. Cambridge,UK: Cambridge University Press.

Alderson, J. C. (2009). Test review. *Language Testing, 26*(4), 621-631.

Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. *Language Testing*, *9*(2), 101–119. doi:10.1177/026553229200900201

Anderson, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.

Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. *The modern language journal, 75*(4), 460-472.

Assiri, M. (2011). *Test-taking strategy use on the reading section of the TOEFL iBT: A study of Arab ESL learners*. PhD dissertation, Oklahoma State University.

Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford,UK: Oxford University Press.

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests*. Oxford,UK: Oxford University Press.

Bachman, L. F., & Palmer, A. (2010). *Language assessment in practice*. Oxford,UK: Oxford University Press.

Bax, S., & Wei, C. (2012). Investigating learners' cognitive reading processes during a computer-based CAE reading test. *University of Cambridge ESOL Examinations Research Note, 47*, 3–14 .

Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, *30*(3), 1–25. doi:10.1177/0265532212473244

Baxter, G. P. and Glaser, R. (1998), Investigating the cognitive complexity of science assessments. *Educational Measurement: Issues and Practice, 17,* 37–45. doi:10.1111/j.1745-3992.1998.tb00627.x

Bernhardt, E. (2005). Progress and procrastination in second language reading. *Annual review of applied linguistics,25,* 133-150.

Bernhardt, E. B. (2011). *Understanding advanced second-language reading.* New York, NY: Routledge.

Bernhardt, E. B., & Kamil, M. L. (1995). Interpreting relationships between L1 and L2 reading: Consolidating the linguistic threshold and the linguistic interdependence hypotheses. *Applied Linguistics, 16*(1), 15-34.

Biber, D., & Gray, B. (2013). *Discourse Characteristics of Writing and Speaking Task Types on the TOEFL iBT® Test: A Lexico-Grammatical Analysis*. ETS Research Report RR-13-04, TOEFLiBT-19. Princeton, NJ: Educational Testing Service.

Bölükbaş, F. (2013). The effect of reading strategies on reading comprehension in teaching Turkish as a foreign language. *Educational Research and Reviews, 8*(21), 2147-2154.

Borsboom, D. (2005).*Measuring the mind: Conceptual issues in contemporary psychometrics.* Cambridge, UK: Cambridge University Press.

Bowers, P.O., Golden, J., Kennedy, A. & Young, A. (1994). Limits upon orthographic knowledge due to processes indexed by naming speed. In: V.W. Berninger (Ed.), *The varieties of orthographic knowledge, Vol. I: Theoretical and developmental issues* (pp. 173–218). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Brantmeier, C. (2005). Effects of reader's knowledge, text type, and test type on L1 and L2 reading comprehension in Spanish.*The Modern Language Journal, 89*(1), 37-53.

Brantmeier, C. (2006). Toward a multicomponent model of interest and l2 reading: sources of interest, perceived situational interest, and comprehension. *Reading in a Foreign Language, 18*(2), 89-115.

Breznitz, Z. (2006). *Fluency in reading: Synchronization of processes*. Mahwah, NJ: Erlbaum

Brown, J. D. (2005). *Testing in language programs: A comprehensive guide to English language assessment* (New ed.). New York, NY: McGraw-Hill.

Bybee, J. (2008). Usage-based grammar and second language acquisition. In P. Robinson & N. Ellis (Eds.), *Handbook of cognitive linguistics and second language acquisition* (pp. 216–236). New York, NY: Routledge

Carr, T. H., Brown, T. L., Vavrus, L. G., & Evans, M. A. (1990). Cognitive skill maps and cognitive skill profiles: Componential analysis of individual differences in children's reading efficiency. In T. H. Carr, & B. Levy (Eds.), *Reading and its development: Component skills approaches* (pp. 1–55). San Diego, CA: Academic Press.

Carr, T. H., & Levy, B. (Eds.). (1990). *Reading and its development: Component skills approaches*. San Diego, CA: Academic Press.

Carrell, P. L., & Eisterhold, J. C. (1983), Schema theory and ESL reading pedagogy. T*ESOL Quarterly*, *17*, 553–573. doi: 10.2307/3586613

Carrell, P. L., & Wise, T.E. (1998). The relationship between prior knowledge and topic interest in second language reading. *Studies in Second Language Acquisition, 20*, 285-309.

Chamot, A. U. (1987). The learning strategies of ESL students. In A. Wenden & J. Rubin (Eds.), *Learner strategies in language learning.* Oxford, UK: Pergamon Press.

Chamot, A., & O'Malley, J. (1994). Instructional approaches and teaching procedures. In R. Spangenberg-Urbschat & R. Pritchard (Eds), *Kids come in all languages: Reading instruction for ESL students* (pp. 82-107). Newark, DE : International Reading Association (IRA).

Chan, S.H.C. (2013) *Establishing the validity of reading-into-writing test tasks for the UK academic context*. PhD dissertation. University of Bedfordshire.

Chapelle, C. A., Enright, M. K., & Jamieson, J. M. (Eds.) (2008). *Building a validity argument for the Test of English as a Foreign Language™*. Mahwah, NJ: Lawrence Erlbaum.

Cho, Y., & Bridgeman, B. (2012). Relationship of TOEFL iBT(R) scores to academic performance: Some evidence from American universities. *Language Testing*, *29*(3), 421–442. doi:10.1177/0265532211430368

Cohen, A.D. (2007). The coming of age for research on test-taking strategies. In J. Fox, M. Weshe, D. Bayliss, L. Cheng, C. Turner, & C. Doe (Eds.), *Language testing reconsidered* (pp. 80-111). Ottawa, Canada: Ottawa University Press

Cohen, A.D. and Upton, T.A. (2006): *Strategies in responding to the New TOEFL reading tasks* (TOEFL Monograph Series Report No. 33). Princeton, NJ: Educational Testing Service. http://www.ets.org/Media/Research/pdf/RR-06-06.pdf

Cohen, A. D., & Upton, T. A. (2007). `I want to go back to the text': Response strategies on the reading subtest of the new TOEFL(R). *Language Testing*, *24*(2), 209–250. doi:10.1177/0265532207076364

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp 443–507). Washington, DC: American Council on Education.

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52*(4), 281-302.

Crossley, S. A., Allen, D. B., & Mcnamara, D. S. (2011). Text readability and intuitive simplification : A comparison of readability formulas. *Reading in a Foreign Language*, *23*(1), 84–101.

Cummins, J. (1990). Empowering minority students: A framework for intervention. *Harvard Educational Review, 56*, 18–36.

Cummins, J. (1991). Interdependence of first-and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70-89), Cambridge, UK: Cambridge University Press.

Cummins, J. (1996). *Negotiating identities: Education for empowerment in a diverse society*. Ontario, CA: California Association for Bilingual Education.

Daneman, M., & Carpenter, P. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*, 450–466 .

Diamond, J. J., & Evans, W. J. (1972). An investigation of the cognitive correlates of test wiseness. *Journal of Educational Measurement, 9*, 145-150.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309-331.

Enright, M. K., Bridgeman, B., Eignor, D., Lee, Y. W., & Powers, D. E. (2008). Prototyping measures of listening, reading, speaking, and writing. In C. Chapelle, M. Enright, & J. Jamieson (Eds.). *Building a validity argument for the Test of English as a Foreign Language* (pp. 145-186). New York,UK: Routledge.

Enright, M. K., & Schedl, M. (2000). *Reading for a reason: Using reader purpose to guide test design*. Unpublished manuscript. Princeton, NY: Educational Testing Service.

Fender, M. (2008). Spelling knowledge and reading development: Insights from Arab ESL learners. *Reading in a Foreign Language, 20*(1), 19-42.

Educational Testing Service. (2009). *The official guide to the TOEFL test* (3rd ed.). New York, NY: McGraw-Hill.

Field, J. (2011). Into the mind of the academic listener. *Journal of English for Specific Purposes, 10,* 102-112.

Field, J. (2012). The cognitive validity of the lecture-based question in the IELTS listening paper. In P. Thompson (Ed.), *IELTS Research Reports* (Vol. 9, pp. 17–66). London, UK: British Council.

Field, J. (2013). Cognitive validity. *Examining Listening: Research and Practice in Assessing Second Language Listening, Studies in Language Testing, 35*, 77-151.

Fitzgerald, J. (1995). English-as-a-second-language learners' cognitive reading processes: A review of research in the United States. *Review of Educational Research*, *65*(2), 145–190. doi:10.3102/00346543065002145

Fraser, C. A. (2007). Reading rate in L1 Mandarin Chinese and L2 English across five reading tasks. *The Modern*, *91*(3), 372–394.

Fuchs, L. S., Fuchs, D., Hosp, M. K., & Jenkins, J. R. (2001). Oral reading fluency as an indicator of reading competence: A theoretical, empirical, and historical analysis. *Scientific studies of reading, 5*(3), 239-256.

Fukkink, R. G., Hulstijn, J., & Simis, A. (2005). Does training in skills affect word reading recognition comprehension? *The Modern Langauge Journal, 89*(1), 54–75.

Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research.* Mahwah, NJ: Lawrence Erlbaum Associates.

Glaser, R. (1981). The future of testing: A research agenda for cognitive psychology and psychometrics. *American Psychologist, 36*, 923-936.

Gomez, P. G., Noah, A., Schedl, M., Wright, C., & Yolkut, A. (2007). Proficiency descriptors based on a scale-anchoring study of the new TOEFL iBT reading test. *Language Testing*, *24*(3), 417–444. doi:10.1177/0265532207077209

Goodman, K. S. (1986).*What's Whole in Whole Language? A Parent/Teacher Guide to Children's Learning*. Portsmouth,NH: Heinemann Educational Books, Inc.

Gorin, J. S. (2006). Test design with cognition in mind. *Educational measurement: Issues and practice, 25*(4), 21-35.

Gorin, J. S. (2007). Reconsidering issues in validity theory. *Educational Researcher, 36*(8), 456-462.

Gottardo, A., & Mueller, J. (2009). Are first- and second-language factors related in predicting second-language reading comprehension? A study of Spanish-speaking children acquiring English as a second language from first to second grade. *Journal of Educational Psychology*, *101*(2), 330–344. doi:10.1037/a0014320

Gough, P.B. (1972). One second of reading. In J.F. Kavanagh & I.G. Mattingly (Eds.), *Language by ear and by eye* (pp. 331–358). Cambridge: MIT Press.

Grabe, W. (2004). Research on teaching reading. *Annual review of applied linguistics*, *24*, 44-69.

Grabe, W. (2010). Fluency in reading — Thirty-five years later. *Reading in a Foreign Language*, *22*(1), 71–83.

Grabe, W., & Stoller, F. (2002). *Teaching and researching reading*. Harlow, UK: Pearson Education.

Green, A., Unaldi, A., & Weir, C. (2010). Empiricism versus connoisseurship: Establishing the appropriacy of texts in tests of academic reading. *Language Testing*, *27*(2), 191–211. doi:10.1177/0265532209349471

Gui, M. (2013). *Relative significanc of component skills to EFL reading: Implications for diagnosis and instruction*. Paper presented at the Langague Teting Research Colloquium (LTRC), Seoul, Korea.

Hamp-Lyons, L. (1997). Washback, impact and validity: Ethical concerns. *Language Testing, 14*(3), 295-303.

Hellekjær, G. O. (2009). Academic English reading proficiency at the university level : A Norwegian case study, *Reading in a Foreign Language, 21*(2), 198–222.

Henning, G. (1987). *A guide to language testing: Development, evaluation, research*. Cambridge, MA: Newbury House.

Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and writing,2*(2), 127-160.

Hsu, W.-T. (2008). *EFL students' English language knowledge, strategy use and multiple-choice reading test performance: A structural equation modeling approach*. Ph.D. Dissertation, University of Southhampton.

Huang, J., & Newbern, C. (2012). The effects of metacognitive reading strategy instruction on reading performance of adult ESL learners with limited english and literacy skills. *Journal of Research and Practice for Adult Literacy, Secondary, and Basic Education*, *1*(2), 66-77.

Inhoff, A. W., & Rayner, K. (1986). Parafoveal word processing during eye fixations in reading: Effects of word frequency. *Perception & Psychophysics, 40*(6), 431-439.

In'nami, Y., & Koizumi, R. (2009). A meta-analysis of test format effects on reading and listening test performance: Focus on multiple-choice and open-ended formats. *Language Testing*, *26*(2), 219–244. doi:10.1177/0265532208101006

Jang, E. E. (2009a). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion Model application to LanguEdge assessment. *Language Testing*, *26*(1), 31–73. doi:10.1177/0265532208097336

Jang, E. E. (2009b). Cognitive diagnostic assessment of L2 reading comprehension ability: Validity arguments for Fusion Model application to LanguEdge assessment. *Language Testing*, *26*(1), 31–73. doi:10.1177/0265532208097336

Jamieson, J. M., Eignor, D., Grabe, W., & Kunnan, A. J. (2008). Frameworks for a new TOEFL. In C. Chapelle, M. Enright, & J. Jamieson (Eds.). *Building a validity argument for the Test of English as a foreign language* (pp. 55-95). New York, NY: Routledge.

Jeon, E. H. (2011). Contribution of morphological awareness to second-language reading comprehension. *The Modern Language Journal*, *95*(2), 217–235. doi:10.1111/j.1540-4781.2011.01179.x

Jeon, E. H. (2012). Oral reading fluency in second language reading. *Reading in a Foreign Language*, *24*(2), 186–208.

Jeon, E. H., & Yamashita, J. (2014). L2 reading comprehension and its correlates: A meta-analysis. *Language Learning*, *64*(1), 160–212. doi:10.1111/lang.12034

Jiang, X., Sawaki, Y., & Sabatini, J. (2012). Word reading efficiency, text reading fluency, and reading comprehension among Chinese learners of English. *Reading Psychology*, *33*(4), 323–349.

Judith, R. (1995). Student responses to reading strategies instruction, *Foreign Language Annals, 28*(2), 262-273.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of educational psychology, 78*(4), 243-255.

Juhasz, B. J., & Rayner, K. (2006). The role of age of acquisition and word frequency in reading: Evidence from eye fixation durations. *Visual Cognition, 13*(7-8), 846-863.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review, 87*, 329-354.

Kalechstein, P. B., Hocevar, D., & Kalechstein, M. (1998). Effects of test-wiseness training on test anxiety, locus of control and reading achievement in elementary school children. *Anxiety Research, 1*(3), 247-261.

Kamhi-Stein, L. D. (2003). Reading in two languages: how attitudes toward home language and beliefs about reading affect the behaviors of "underprepared" L2 college readers. *TESOL Quarterly, 37*(1), 35-71.

Karami, S., & Hashemian, M. (2012). The relationship between (meta)cognitive strategies and reading comprehension in iranian female l2 learners. *International Journal of English Linguistics*, *2*(4), 58–64. doi:10.5539/ijel.v2n4p58

Khalifa, H., & Weir, C. J. (2009). *Examining Reading: Research and practice in assessing second language reading*. Cambridge University Press

Killen, R. (2003). Validity in outcomes-based assessment. *Perspectives in Education, 21*(1), 1-14.

Kintsch, W. (1998). *Comprehension: A paradigm for cognition.* Cambridge, UK: Cambridge University Press.

Koda, K. (1996). L2 word recognition research: A critical review. *The Modern Language Journal, 80*(4), 450-460.

Koda, K. (2005). *Insights into second language reading: A cross-linguistic approach*. Cambridge, UK: Cambridge University Press.

Koda, K. (2007). Reading and language learning: Crosslinguistic constraints on second language reading development. *Language Learning, 57*(1), 1-44.

Lado, R. (1961). *Language Testing*: New York, NY: McGraw-Hill.

Larson-Hall, J. (2010). *A guide to doing statistics in second language research using SPSS*. New York, NY: Routledge.

Lee, Y., & Sawaki, Y. (2009). Cognitive diagnosis approaches to language assessment: An overview. *Language Assessment Quarterly*, *6*(3), 37–41. doi:10.1080/15434300902985108

Leeser, M. J. (2007). Learner-based factors in L2 reading comprehension and processing grammatical form: Topic familiarity and working memory. *Language Learning, 57*(2), 229-270.

Lemaire, B., Guerin-Dugue, A., Baccino, T., Chanceaux, M., & Pasqualotti, L. (2011). A cognitive computational model of eye movements investigating visual strategies on textual

material. In L. Carlson, C. Hölscher, & T. Shipley (Eds.). *Proceedings of the Annual Meeting of the Cognitive Science Society* (pp. 1146–1151). Boston, Massachusetts: Cognitive Science Socieity.

Lim, H., & Godfroid, A. (2014). Automatization in second language sentence processing: A partial, conceptual replication of Hulstijn, Van Gelderen, and Schoonen's 2009 study. *Applied Psycholinguistics*, Advance online publication. http://dx.doi.org/10.1017/S0142716414000137

Lin, L.C., & Yu, W.Y. (2013). A think-aloud study of strategy use by EFL college readers reading Chinese and English texts. *Journal of Research in Reading*, Advance online publication. doi:10.1111/1467-9817.12012

Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, *34*(4), 207–218.

Matin, E. (1974). Saccadic suppression: A review and an analysis. *Psychological Bulletin, 81*(12), 899–917.

McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of reading, 12*(8), 639-646.

Melby-Lervåg, M., & Lervåg, A. (2013). Reading comprehension and its underlying components in second-language learners: A meta-analysis of studies comparing first- and second-language learners. *Psychological Bulletin, 140*(2), 409-433. doi:10.1037/a0033890

Messick, S. (1980). Test validity and the ethics of assessment. *American Psychologist, 35*(11), 1012–1027.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 13-103). New York, NY: American Council on Education & Macmillan.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. *Educational Researcher, 23*(8), 13-23.

Messick, S. (1995). Standards of validity and the validity of standards in performance assessment. *Educational Measurement: Issues and Practice, 14*(4), 5-8.

Messick, S. (1996). Validity and washback in language testing. L*anguage Testing, 13*(3), 241-256.

Miller, B., & O'Donnell, C. (2013). Opening a window into reading development: Eye movements' role within a broader literacy research framework. *School Psychology Review, 42*(2), 123-139.

Millman, J., Bishop, C. H., & Ebel, R. (1965). An analysis of test-wiseness. *Educational & Psychological Measurement, 25*(3), 707-726.

Mitchell, D. C., & Green, D. W. (1978). The effects of context and content on immediate processing in reading. *The Quarterly Journal of Experimental Psychology, 30*(4), 609-636.

Morrison, R. E. (1984). Manipulation of stimulus onset delay in reading: Evidence for parallel programming of saccades. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 667-682.

Naseri, M., & Zaferanieh, E. (2012). The relationship between reading self-efficacy beliefs, reading strategy use and reading comprehension level of Iranian EFL learners. *World Journal of Education*, *2*(2), 64–75. doi:10.5430/wje.v2n2p64

Nassaji, H. (2003). Higher–level and lower–level text processing skills in advanced esl reading comprehension. *The Modern Language Journal, 87*(2), 261-276.

Nassaji, H. (2014). The role and importance of lower-level processes in second language reading. *Language Teaching, 47*(1), 1-37.

Nguyen, L. T. C., & Nation, I. S. P. (2011). A bilingual vocabulary size test of English for Vietnamese learners. *RELC Journal, 42*(1), 86-99.

Norris, J. M. (2005). Using developmental sequences to estimate ability with english grammar : preliminary design and investigation of a web-based test . *Second Language Studies*, *24*(1), 24–128.

Overmann, M. (1999), Reading is guided creation*. Fremdsprachenunterricht, 43*(5) 327-332.

Paulson, E. J. (2000). *Adult readers' eye movements during the production of oral miscues.* Unpublished doctoral dissertation. Tucson, AZ: University of Arizona..

Paulson, E. J., & Henry, J. (2002). Does the degrees of reading power assessment reflect the reading process ? An eye-movement examination. *Journal of Adolescent & Adult Literacy*, *46*(3), 234–244.

Perfetti, C.A. (1985). *Reading ability.* New York,NY: Oxford Press.

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, *20*(1), 25–56.

Phillips, N., Segalowitz, N., O'Brien, I., & Yamasaki, N. (2004). Semantic priming in a first and second language: Evidence from reaction time variability and event-related brain potentials. *Journal of Neurolinguistics*, *17*(2-3), 237–262. doi:10.1016/S0911-6044(03)00055-1

Pollatsek, A., Reichle, E. D., & Rayner, K. (2006d). Tests of the E-Z Reader model: Exploring the interface between cognition and eye-movement control. *Cognitive Psychology, 52*(1), 1–56.

Pressley, M. (1998). Comprehension strategies instruction. In J. Osborn, & F. Lehr (Eds.), *Literacy for all: Issues in teaching and learning* (pp. 113–133). New York,NY: Guilford

Prince, P. (2014). Listening comprehension: Processing demands and assessment issues. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.). *Measuring L2 proficiency: Perspectives from SLA* (pp. 93-108), Bristol,UK: Multilingual matters.

Pulido, D. (2004). The relationship between text comprehension and second language incidental vocabulary acquisition: A matter of topic familiarity? *Language learning, 54*(3), 469-523.

Purpura, J. E. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language learning, 47*(2), 289-325.

Purpura, J. E. (1998). Investigating the effects of strategy use and second language test performance with high- and low-ability test takers: A structural equation modelling approach. *Language Testing*, *15*(3), 333–379. doi:10.1177/026553229801500303

Purpura, J. E. (2004). *Assessing grammar.* Cambridge,UK: Cambridge University Press.

Qian, D. D., & Schedl, M. (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing*, *21*(1), 28–52. doi:10.1191/0265532204lt273oa

Rayner, K. (1998). Eve movements in reading and information processing: 20 years of research . *Psychological Bulletin, 124*(3), 372 – 422.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology*, *62*(8), 457–506. doi:10.1080/17470210902816461

Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, *10*(3), 241–255.

Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs, NJ: Prentice Hall

Rayner K., Pollatsek A., Ashby J., Clifton C., Jr. (2012). *Psychology of Reading*. 2nd ed. New York, NY: Psychology Press

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*(3), 164–171. doi:10.1027/1015-5759/a000123

Reed, D., Ohlrogge, A., Lim, H., & Kim, H. (2013, July). *Complexity comparisons between examination passages and real-world reading texts at 2 CEFR levels*. Poster presented at the Langague Teting Research Colloquium (LTRC), Seoul, Korea.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner,K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*(1), 125–157.

Reichle, E. D., Warren, T., & McConnell, K. (2009). Using E-Z Reader to model the effects of higher level language processing on eye movements during reading. *Psychonomic Bulletin & Review, 16*(1), 1–21.

Roever, C., & Powers, D. E. (2005). *Effects of language of administration on a self-assessment of language skills.* Princeton, NJ: Educational Testing Service.

Rogers, W. T., & Bateson, D. J. (1991). The influence of test-wiseness on performance of high school seniors on school leaving examination. *Applied Measurement in Education, 4*(2), 159-183.

Rogers, W. T., & Harley, D. (1999). An empirical comparison of three-and four-choice items and tests: Susceptibility to testwiseness and internal consistency reliability. *Educational and Psychological Measurement*, *59*(2),234-247.

Rupp, A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, *23*(4), 441–474. doi:10.1191/0265532206lt337oa

Samuels, S. J. (1977). Introduction to theretical models of reading. In W. Otto (Ed), *Reading Problems* (pp. 7-41). Boston, MA: Addison-Wesley.

Sarnaki, R. E. (1979). An examination of test-wiseness in the cognitive domain. *Review of Educational Research, 49*(2), 252-279.

Sawaki, Y., Stricker, L. J., & Oranje, A. H. (2009). Factor structure of the TOEFL internet-based test. *Language Testing, 26*(1), 5-30.

Schmitt, N., Schmitt, D., & Clapham, C. (2001). Developing and exploring the behaviour of two new versions of the Vocabulary Levels Test. *Language testing, 18*(1), 55-88.

Segalowitz, N., & Segalowitz, S. J. (1993). Skilled performance, practice and the differentiation of speedup from automatization effects. *Applied Linguistics, 14*(3), 369–385

Segalowitz, N., & de Almeida, R. G. (2002). Conceptual representation of verbs in bilinguals: Semantic field effects and a second-language performance paradox. *Brain and Language*, *81*(1-3), 517–531. doi:10.1006/brln.2001.2544

Segalowitz, N., & Frenkiel-Fishman, S. (2005). Attention control and ability level in a complex cognitive skill: Attention shifting and second-language proficiency. *Memory & Cognition*, *33*(4), 644–53. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/16248329

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in second language learning. In J. F. Kroll, & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). Oxford, UK: Oxford University Press.

Sheorey, R., & Mokhtari, K. (2001). Differences in the metacognitive awareness of reading strategies among native and non-native readers. *System*, *29*(4), 431–449. doi:10.1016/S0346-251X(01)00039-2

Shiotsu, T. (2003): *Linguistic knowledge and processing efficency as predictors of L2 reading ability: a component skills analysis*. Unpublished PhD thesis, The University of Reading.

Shiotsu, T., & Weir, C. J. (2007). The relative significance of syntactic knowledge and vocabulary breadth in the prediction of reading comprehension test performance. *Language Testing*, *24*(1), 99–128. doi:10.1177/0265532207071513

Solheim, O. J. (2011). The impact of reading self- efficacy and task value on reading comprehension scores in different item formats. *Reading Psychology*, *32*(1), 1–27.

Solheim, O. J., & Uppstad, P. H. (2011). Eye-tracking as a tool in process-oriented reading test validation. *International Electronic Journal of Elementary Education*, *4*(1), 153–168.

Song, X., & Cheng, L. (2009). Language learner strategy use and test performance of Chinese learners of English. *Language Assessment Quarterly*, *3*(3), 37–41.

Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading research quarterly*, *16*(1), 32-71.

Stanovich, K. E. (1984). The interactive-compensatory model of reading: A confluence of developmental, experimental, and educational psychology. *Remedial and Special Education*, *5*(3), 11–19. doi:10.1177/074193258400500306

Stanovich, K. E. (1988). Explaining the differences between the dyslexic and the garden-variety poor reader: The phonological-core variable-difference model. *Journal of learning disabilities,21*(10), 590-604.

Stanovich, K. E. (2000). *Progress in understanding reading: Scientific foundations and new frontiers.* New York,NY: Guilford Press.

Storey, P. (1997). Examining the test-taking process: A cognitive perspective on the discourse cloze test. *Language Testing*, *14*(2), 214–231. doi:10.1177/026553229701400205

Stricker, L. J. (2004). The performance of native speakers of English and ESL speakers on the computer-based TOEFL and GRE General Test. *Language Testing*, *21*(2), 146–173. doi:10.1191/0265532204lt279oa

Stricker, L. J., & Attali, Y. (2010). *Test takers' attitudes about the TOEFL iBT [TM]*. TOEFL iBT Research Report. RR-10-2. Princeton, NJ: Educational Testing Service.

Stricker, L. J., & Rock, D. A. (2008). *Factor structure of the TOEFL Internet-based test across subgroups*. TOEFL iBT Research Report No. TOEFLiBT-07. Princeton, NJ: Educational Testing Service.

Taillefer, G., & Pugh, T. (1998). Strategies for professional reading in L1 and L2. *Journal of Research in Reading, 21*(2), 96-108.

Thelk, A. D., & Hoole, E. R. (2006). What are you thinking? Postsecondary student think-alouds of scientific and quantitative reasoning items. *The Journal of General Education, 55*(1), 17-39.

Tsai, Y. R., Ernst, C., & Talley, P. C. (2010). L1 and L2 strategy use in reading comprehension of Chinese EFL readers. *Reading Psychology*, *31*(1), 1-29.

Urquhart, S., & Weir, C. (1998). *Reading in a second language: Process, product and practice*. London, UK: Longman.

Van Gelderen, A., Schoonen, R., de Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2004). Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: A componential analysis. *Journal of Educational Psychology*, *96*(1), 19–30. doi:10.1037/0022-0663.96.1.19

Van Gelderen, A., Schoonen, R., Stoel, R. D., De Glopper, K., & Hulstijn, J. (2007). Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology, 99*(3), 477-491.

Van Keer, H., & Verhaeghe, J. P. (2005). Effects of explicit reading strategies instruction and peer tutoring on second and fifth graders' reading comprehension and self-efficacy perceptions. *The Journal of Experimental Education, 73*(4), 291-329.

Verhoeven, L. T. (2000). Components in early second language reading and spelling. *Scientific Studies of Reading, 4,* 313–330.

Von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, *65*(2), 109–127. doi:10.1016/j.jml.2011.02.004

Walter, C. (2007). First- to second-language reading comprehension: Not transfer, but access. *International Journal of Applied Linguistics, 17*(1)*,* 14–37. doi:10.1111/j.1473-4192.2007.00131.

Weir, C. J. (2005). *Language testing and validation*. New York, NY: Palgrave MacMillan.

Weir, C. J., & Hawkey, R. G. A., & Devi, S.(2009a). The cognitive processes underlying the academic reading construct as measured by IELTS.*IELTS Research Reports, 9,* 157-189.

Weir, C., Hawkey, R., Green, A., Unaldi, A., & Devi, S. (2009b). The relationship between the academic reading construct as measured by IELTS and the reading experiences of students in their first year of study at a British university. *IELTS Research Reports, 9*, 97-156.

Wickens (1992). *Engineering psychology and human performance* (2nd ed.). New York, NY: Harper Collins.

Williams, E., & Moran, C. (1989). Reading in a foreign language at intermediate and advanced levels with particular reference to English. *Language Teaching, 22*(4), 217-228.

Winke, P. (2011). Evaluating the validity of a high-stakes ESL test: Why teachers' perceptions matter. *TESOL Quarterly, 45*(4), 628-660.

Yamashita, J. (2003). Processes of taking a gap-filling test: Comparison of skilled and less skilled EFL readers. *Language Testing*, *20*(3), 267–293. doi:10.1191/0265532203lt257oa

Yang, P. (2000). *Effects of test-wiseness upon performance on the test of English as a foreign language.* Ph.D. dissertation, University of Alberta, Edmonton, Alberta, Canada. Available at http://www.collectionscanada.gc.ca/

Yeh, L. H., Schwartz, A. I., & Baule, A. L. (2011). The impact of text-structure strategy instruction on the text recall and eye-movement patterns of second language English readers. *Reading Psychology, 32*(6), 495-519.