**Integrated Language Skills and Its Assessment: A Literature Review**

**ABSTRACT**

The paper looked at the theoretical premise and practical implementations of assessment for integrated language skills for English as a second language. It traced back the evolution of assessment of integrated language skills from isolated skills to integration on the premise of constructivist theory while reflecting and working on real world examples. This main objective of this paper was to explore and critically review integrated language skills and its assessment process. To meet this objective the researcher explored various researches conducted and papers written over the years. This paper investigated historical perspectives and progression, recent approaches and researches pertaining to assessment of integrated language skills. It further accentuated the fact that these approaches are aligned with recent models of multiliteracy and lay out a better picture of academic language competence. Key benefits like ameliorated authenticity, enhanced validity, positive backwash effect, relevant diagnostic information etc. were identified in the paper after analysing various studies. On one hand the research acknowledged challenges of implementing in relation to swift development, rating and evaluation system, construct definition; on the other hand, it inferred that assessment of integrated skills marks noteworthy improvement in the methodology of language testing. The findings underpinned pedagogical approaches that foreground integrated language skills as ‘inevitably vital’ for the progression and development of communicative competence among learners.

**Keywords:** Integrated skills assessment, language testing, constructivist theory, discourse synthesis, task-based instruction.

**1. INTRODUCTION**

According to the constructivist theory, learners learn language and construct knowledge through the experiences that they gain via interaction and reflections (Vygotsky, 1978, 1979). For this to take effect, learners’ autonomy needs grasp over one’s learning needs and strategies, while teachers act as the resource and facilitator to meet those needs (Voller, 1997, cited in Benson Anduy Voller, 1997). English language learners face various problems in the academic sphere, including unfulfilled expectations, culture shock and difficulty in adjustment, academic problems. Krashen (1987) suggested that lowers performance among adults may be the because of high anxiety levels, thus proposing ‘comprehensive input’ as necessary for the learning of second language (Krashen, 1981, 1982, 1985). Harley (1986) observed that natural environment for language learning produces better results (Brown, 1991a; Dulay, Burt & Krashen, 1982 Gass & Selinker, 2001; Krashen, 1987). The four language skills i.e., speaking, writing, reading and listening have been conceptualised as ‘macro-skills’ by language educators. Reading and listening are considered as receptive skills, while speaking and writing are productive skills. These macro skills when taught in integrated manner allows the learner to cultivate their language skills in an efficient and effective manner. Though the assessment of these integrated language skills may require a touch of innovation from the teacher. According to Temple and Gillet (1984), there is a deep relationship between speaking and listening. Reading as a receptive skill is usually developed alongside speaking and listening. However, writing requires structured presentation of thoughts and is beyond mere transcription of speech; thus, earning its title as one of the most challenging language skills. Whole-language theorists advocates for integrated language skills opportunities in meaningful activities, since all language skills aspects are interrelated (Carrasquillo, 1993; Farris, 1989; Farris and Kaczmarski, 1988).

Richards (1983, cited in Omaggio, 1986) discusses micro skills involved in listening includes discrimination of sound, recognition of word boundary and identification of syntactic pattern. Abdel-Salam El-Koumy (2002) mentions how language is viewed as separate skills and is taught in predetermined sequence, while the skills-based approach stems from behavioural psychology and structural linguistics. A correlation has been indicated between vocabulary knowledge and reading comprehension through research (Barr, 1985; Hoover and Gough, 1990; Kitao, 1988), with improved vocabulary reading comprehension enhances as well (Cziko, 1980; Davis, 1989; McDaniel and Pressley, 1986). Nation (1989) recommends moving forward from individual to collaborative learning activities, while Ibrahim (1993) supports an integrated approach that combines linguistic skills as well as communicative abilities. Before 19600, language skills were taught in segregated manner however, for the holistic development of language, integrated language skills came into the spotlight. To ensure language skills are nurtured in integrated manner the need to assess them was a dimension worth exploring.

**2. HISTORY**

Lado mentions how criticism of conventional methods of assessment emerged in the 1970s when educators started supporting the evaluation of communicative competence instead of grammar and vocabulary in isolation (Lado, 1961).

While these efforts expanded the range of competencies tested, they still maintained Carroll's (1975) influential distinction of four separate language "skills" that became standard in educational curricula worldwide. This demarcation of categories has been challenged based on numerous reasons. When applied to something as complicated as communication, the term ‘skill’ proves itself to be problematic as indicated by research conducted (Anderson, 1995). Koda (2007) discusses that just reading alone exhibits multiple subskills that are interrelated, instead of it being a single ability.

During the use of language, it in itself integrates multiple abilities and skills, since people discuss on a regular basis about what they have read or heard. Pertaining to the context of education, conventional assessments done in the form of writing test papers has been criticised by educators for being lacking since it tests limited aspects of language skills (Morrow, 1977; Lewkowicz, 2000), which in turn influences teachers negatively to practice formulaic writing types (Raimes, 1990). These observations exhibit how tests tend to under-represent the construct of writing for academic purposes, a concern that has intensified as formal testing has gained momentum in educational policies (Hillocks, 2002). Similar concerns were raised regarding the assessment of speaking skills using orally conducted tests and interviews (van Lier, 1989). According to Pierce (1992), the language tests with test-specific genres and areas assessed do facilitate data regarding the measurement of the skills; however, they poorly accommodate and represent the use of language in real world.

In response, performance assessments were developed to realistically represent complex communication types found in university and workplace settings. The author traces how Morrow's (1977) influential conceptualization of communication assessment led to innovative integrated language tests for the Royal Society of Arts (Hawkey, 2004), while Davies (2008) documented the English Language Testing System's attempts to systematically integrate academic language skills. Wesche (1987) provided a notable example with the Ontario Test of ESL, which mirroring required examinees to write and speak critically about source texts they had interpreted, authentic university course assignments.

**3. CONTEMPORARY VIEWS**

In recent times, academicians discuss the shift toward interactionist conceptualizations in language assessment. Unlike traditional approaches that view language ability as fixed traits independent of context, interactionist assessments evaluate how examinees perform specific complex tasks representing crucial activities in real language that uses contexts (Mislevy & Yin, 2009). This aligns with recent educational trends promoting multiliteracies rather than autonomous skills (Cope & Kalantazis, 2000).

Integrated skills assessments embrace constructivist principles of knowledge integration as theorized by Kintsch (1998) and Bereiter (2002). They typically involve tasks requiring examinees to interpret source information and produce written or spoken responses for specific purposes, which are then rated based on performance criteria. The TOEFL exemplifies this approach with integrated writing tasks that simulate university exam conditions.

Research justifies these assessments for higher education admissions by demonstrating that university writing primarily involves displaying knowledge from sources in appropriate academic conventions (Leki & Carson, 1997; Sternglass, 1997; Rosenfeld, Leung, & Oltman, 2001; Leki, 2007; Byrnes, 2008). Similar integrated assessments exist globally, including Canada's CAEL Assessment and New Zealand's DELNA.

Another approach identifies a spectrum of integrated skills assessments: weaker versions like Cambridge's First Certificate in English requiring minimal integration; middle-range versions like TOEFL requiring synthesis of lengthy materials; and stronger versions in task-based assessments representing authentic communication tasks determined through needs analyses (Norris, 2002; Hawkey, 2004; Colpin & Gysen, 2006; Deane, 2011). These approaches reflect the growing recognition that successful language use in academic contexts requires the integration of multiple language skills rather than isolated competencies.

**4. RECENT RESEARCHES**

Plakans’ (2012) concept of integrated skills assessment in language testing, compares it with traditional single-skill assessments. It distinguishes three types of integrated assessments: content-responsible tasks requiring test-takers to comprehend and incorporate source materials, stimulus-related tasks using texts as idea generators without requiring their integration, and thematically linked assessments connecting multiple test sections by topic. Research by Cumming et al. (2005) found integrated writing responses were typically shorter but used more diverse vocabulary than independent tasks, while Watanabe (2001) observed that integrated essays were less likely to contain original theses. Studies on test-taking processes reveal that some test-takers approach both task types similarly, while others employ specific integration strategies for integrated tasks (Plakans, 2008, 2009). Ascención (2005) discovered differences in planning requirements between various integrated task types. Research also indicates that performance on integrated tasks varies with proficiency level, prior experience, educational background, and first-language writing experience (Ruiz-Funes, 2001; Ascención, 2005; Plakans, 2009).

Cumming (2024) talks about recent research and its trends in assessment of integrated language skills. Cumming's (2013) review identifies five key benefits of integrated writing assessments: they provide realistic literacy activities, engage examinees in content-responsible writing, counter test method effects, align with modern literacy models, and offer diagnostic value. Research supporting these claims falls into four approaches. Studies examining examinee processes during integrated tasks have demonstrated these assessments elicit diverse interpretive and analytical strategies that surpass conventional tasks, approximate academic demands, and are perceived as more authentic by students (Cumming, Rebuffot, & Ledwell, 1989; Esmaeili, 2002; Fraser, 2002; Plakans, 2008; Plakans & Gebril, 2012; Yang & Plakans, 2012; Swain, Huang, Barkaoui, Brooks, & Lapkin, 2009).

Discourse analysis research by Cumming et al. (2005), Knoch (2009), Plakans (2009), and Yu (2009) has revealed that compositions from integrated tasks display more complex lexical, syntactic, and rhetorical features compared to compositions without source materials. Frost, Elder and Wigglesworth (2012) found that content quality in speaking tasks correlates with proficiency scores. Studies of instructor perceptions by Cumming, Grant, Mulcahy-Ernt, and Powers (2004), Brown, Iwashita, and McNamara (2005), Wall and Horak (2008), and Knoch (2009) indicate positive impressions of integrated tasks as authentic representations of academic abilities. Research on diagnostic value has examined diverse contexts including DELNA (Knoch, 2009), TOEFL (Sawaki, Quinlan, & Lee, 2013), and engineering writing (Artemeva & Fox, 2010).

Aydoğan, H., & Akbarov, A. A. (2014) discussed the results of their research in their paper. The research presents five key statistical findings about the Quality of Language Experience (QLE) subscales. First, all QLE subscales demonstrate statistically significant correlations with each other. Second, five aspects of language learning (improving general English, vocabulary acquisition, grammar study, reading, and writing) collectively explain 58.8% of variance in speaking English outside the classroom, with general English improvement and writing skills emerging as statistically significant predictors of speaking practice outside class contexts. Third, students have developed more effective strategies for vocabulary learning compared to grammar study. Fourth, participants demonstrate stronger skills in writing than in reading. Fifth, statistically significant gender differences appear only in general English improvement, with other variables showing similar results across genders.

Students who participated in the research recommended extending interactive practice beyond English courses to all subject areas, noting that teacher-student and student-student interaction enhanced interest, affection, and motivation. The research identified writing as the dominant skill emphasized by students in both teaching and assessment compared to other language skills. While no statistically significant differences were found in speaking and reading comprehension for both daily and academic language purposes, the researvchers emphasize the critical importance of regular assessment of written language skills and the development of training programs to enhance these skills.

The researchers conclude that teachers should encourage independent learning by improving instruction quality and moving away from monotonous activities. They advocate for skill integration as "inevitably vital," referencing Rebecca Oxford's metaphor of language class as a tapestry with interconnected strands of language skills (Oxford 2001). The integrated-skill approach exposes learners to authentic language and encourages natural interaction, emphasizing that while teaching separated skills merely teaches language, integrating skills teaches authentic communication (Oxford 2001).

Yang & Plakans (2012) in their research paper examines integrated writing tasks in language assessment, particularly focusing on their construct validity. These tasks, which combine reading, listening, and writing skills, have been incorporated into various language tests since the 1980s, including the TOEFL in recent years. Their inclusion has generated both support and concern among language testing experts (Feak & Dobson, 1996; Read, 1990).

Proponents argue that integrated tasks enhance test fairness by providing content knowledge to all test-takers and promote positive washback effects because they resemble actual academic tasks (Cumming, Grant, Mulcahy-Ernt, & Powers, 2004; Fox, 2004). However, critics point to construct-related validity issues stemming from their multidimensional nature (Charge & Taylor, 1997; Fox, 2003; Upshur & Turner, 1999). As Bachman and Palmer (1996) emphasized, "unless we can demonstrate that the inferences, we make on the basis of language tests are valid, we have no justification for using test scores for making decisions about individuals" (p. 95).

While researchers have examined both the written products and the processes/strategies employed in integrated reading-writing tasks, less is known about reading-listening-writing tasks. This study aims to fill this gap by investigating L2 writers' strategy use, test performance, and the relationship between the two in integrated reading-listening-writing contexts.

The shift toward performance-based writing assessments has occurred because they allow more direct inferences about test constructs from examinees' performances (Messick, 1994). However, conventional writing assessments that test writing in isolation don't accurately reflect real academic writing, which typically integrates multiple skills. Integrated tasks may improve predictive validity (Lewkowicz, 1997; Wesche, 1987) and have positive washback effects (Cumming et al., 2004; Weigle, 2004).

Concerns about integrated tasks include the potential confusion between assessing reading and writing abilities (Charge & Taylor, 1997; Fox, 2003) and verbatim source use issues (Lewkowicz, 1994). To develop a validity argument for these tasks, researchers suggest investigating test-takers' processes and strategies, which are considered important sources of variation in language test performance (Anderson, 1991; Bachman, 1990; Cohen & Upton, 2007).

**5. INTEGRATED LANGUAGE PROCESS WITH SPECIAL FOCUS ON WRITING SKILL**

Regarding integrated language processes, L1 composition theories provide useful frameworks. Spivey and King (1989) and Spivey (1984, 1990, 1997) identified three major discourse synthesis processes: organizing (approaching texts based on prior knowledge of text organization), selecting (sorting key ideas based on task goals), and connecting (linking information from different sources). L2 integrated writing research has added two more operations: monitoring (identifying goals and creating strategic plans) and evaluating (re-examining task effectiveness).

Test-wiseness strategies represent another category, including tricks examinees use to arrive at correct answers rather than engaging in expected cognitive operations. These can create construct-irrelevant variance (Crocker, 2006; Haladyna & Downing, 2004) that decreases score precision. Examples include using memorized writing models, verbatim copying from source materials (Braine, 2001; Cumming et al., 2005; Lewkowicz, 1994; Watanabe, 2001), and patchwriting (Howard, 1995; Hyland, 2005).

Integrated writing tests typically assess content relevance/adequacy, essay organization, and language accuracy. Research by Cumming et al. (2005) found that effective examinees tend to summarize rather than paraphrase source materials, state main positions clearly, and avoid verbatim copying. Less effective examinees used verbatim phrases, failed to cite sources, and often wrote from personal experience without addressing prompts.

This study follows Bachman's (2002) suggestions for understanding test constructs by collecting data on test-takers' responses, their strategies, and using structural equation modelling to analyse the relationship between strategy use and performance. The study involved 161 non-native-English-speaking university students from diverse countries, educational backgrounds, and proficiency levels.

The researchers developed the Strategy Inventory for Integrated Writing (SIIW) to capture test-takers' mental and behavioural activities during integrated writing tasks. After validation through expert review, pilot testing, and factor analysis, the final inventory contained 24 items across six factors: monitoring, test-wiseness, organizing, connecting, evaluating, and selecting. These factors were then analysed as indicators of three latent factors: discourse synthesis strategies, self-regulatory strategies, and test-wiseness strategies. This comprehensive approach allows for a deeper understanding of the construct validity of integrated reading-listening-writing tasks in language assessment.

The study concludes that integrated speaking-reading-listening-writing tasks involve more than just text comprehension and production abilities. L2 writers must activate self-regulatory mechanisms to coordinate various strategies for successful task completion. The findings support the construct validity of these tasks by demonstrating that test-takers engage in the cognitive processes hypothesized in integrated writing literature (Esmaeili, 2002; Plakans, 2009a; Spivey & King, 1989).

Theoretically, the study identifies close linkages between process and product in integrated writing tasks and addresses connections among reading, listening, and writing. It provides evidence for substantive validity (Messick, 1989) by showing that test-takers engage in strategies proposed in integrated writing literature. The positive effect of discourse synthesis strategies and negative effect of test-wiseness strategies on performance help address concerns about construct-irrelevant variance.

The study also found that L2 writing ability was not necessarily proportionate to source use ability, supporting research suggesting that inappropriate source use may stem from various factors including task characteristics, cultural differences, and personal beliefs (Howard, 1995; Kirkland & Saunders, 1991; Pennycook, 1996).

Pedagogically, the findings support an integrated-skills approach to language teaching (Nunan, 1989; Oxford, 2001; Scarcella & Oxford, 1992) and suggest explicit strategy instruction for integrated tasks. Teachers should help students develop "writerly reading strategies" (Hirvela, 2004) and address plagiarism issues while teaching appropriate referencing skills.

The researchers acknowledge several limitations, including differences between the research context and actual high-stakes testing environments, the limited scope of strategy items in the inventory, potential topic effects from using a single task, the restricted sample size of 161 students from one U.S. university, and the possibility that other models might better explain the data. Despite these limitations, the study provides valuable insights into the complex nature of integrated writing tasks and the strategies that contribute to successful performance.

**6. HOW TO ASSESS TASK- BASED** **INSTRUCTION**

The researches suggests that English teachers should create materials addressing student needs while continuously evaluating how effectively they integrate skills. For example, writing instruction should involve drafting and revision processes with student participation and needs analysis rather than conventional timed essays. The study results indicate that integrated-skill approaches, whether content-based or task-based, motivate students of all ages and backgrounds. In task-based instruction, activities require authentic language use with primary focus on meaning rather than form (Nunan, 1989), promoting collaboration among students. The researchers also advocate for the theme-based model (Scarcella & Oxford, 1992), which integrates language skills around engaging themes that facilitate practice of various language skills.

**7. BENEFITS**

The benefits of integrated assessment include authenticity, as they mirror real-world language use contexts where skills are employed in combination rather than isolation (Leki & Carson, 1997). This authenticity potentially increases test-taker motivation and yields scores more predictive of future performance. Additionally, integrated assessments provide content for test-taker responses, potentially mitigating non-language factors like creativity or background knowledge (Read, 1990). Some studies indicate test-takers prefer integrated tasks due to better topic comprehension (Plakans, 2009; Huang, 2010), though Huang (2010) found no evidence of reduced anxiety in actual performance measures. These assessments align well with pedagogical approaches like task-based language teaching and may positively influence classroom practices.

**8. CHALLENGES**

Despite these advantages, integrated assessments present significant challenges. Developing high-quality prompts requires careful consideration of text selection, length, and complexity. Clear instructions are crucial, as test-takers approach these tasks differently. Rating systems must effectively capture skill integration, and assessors must address issues like verbatim copying from source texts (Cumming et al., 2005; Gebril & Plakans, 2009). Validity considerations are particularly complex, with research suggesting integrated tasks measure unique constructs beyond the simple combination of individual skills (Delaney, 2008; Plakans, 2009; Yang, 2009). While integrated skills assessment represents an important evolution in language testing, continued research is necessary to define these constructs more precisely and understand how to measure them effectively.

**9. CONCLUSION**

Language assessment has evolved from testing of skills in isolated form towards integrated approach. With this, one can see a very significant Shift in paradigm towards language education and its evaluation. After many years of research and practise one may safely say that integrated language skills assessment has transpired as valid representation of language used in real world where communication is comprised of various language ability and skills. The constructivist theory supporting this approach agrees with the fact that, through meaningful interaction and experiences, knowledge is built instead of acquiring skills in fragmented manner. Through various researches it has been demonstrated that integrated assessments offer several advantages. It allows learners to be engaged in content specific reading, speaking, writing and listening and acts as a counter for traditional testing. They correspond to the contemporary multiliteracy models while offering valuable insights regarding learners’ abilities.

Studies examining test-taker processes reveal that integrated tasks elicit complex interpretive and analytical strategies that more closely approximate authentic academic demands. However, implementing integrated assessment presents challenges, including developing appropriate prompts, creating clear instructions, establishing valid rating systems, addressing verbatim copying issues, and more precisely defining the constructs being measured. Moving forward, language educators should embrace skill integration as "inevitably vital" while continuing research to refine assessment practices. The integrated-skill approach, whether content-based or task-based, not only enhances assessment authenticity but also positively influences classroom teaching by encouraging meaningful communication over isolated skill practice. As language education continues to evolve, the integration of skills in both teaching and assessment will remain central to developing truly communicative language competence that prepares learners for real-world language use contexts.

**Disclaimer (Artificial intelligence)**

I, Arshi Alam and Prof. Arshad Ikram Ahmad hereby declares that NO generative AI technologies such as Large Language Models (ChatGPT, COPILOT, etc.) and text-to-image generators have been used during the writing or editing of this manuscript.

**Reference**

Abdel-Salam Abdel-Khalek El-Koumy (2002). Teaching and Learning English as a Foreign Language: A Comprehensive Approach, p.12. Suez Canal University, Egypt.

Ablex Publishing Company. Chomsky, N. (1965). Aspects of theory syntax. Cambridge, MA: The M.I.T. Press.

Alexander, P. A., Graham, S., & Harris, K. R. A. (1998). A perspective on strategy research: Progress and prospects. Educational Psychology Review, 10, 129–154. doi:10.1023/A:1022185502996

Allan, A. (1992). Development and validation of a scale to measure test-wiseness in EFL/ESL reading test takers. Language Testing, 9, 101–122. doi:10.1177/ 026553229200900201

Anderson, J. R. (1995). Learning and memory. New York, NY: John Wiley.

Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. Modern Language Journal, 75, 460–472. doi:10.1111/j.1540- 4781.1991.tb05384.x

Anderson, N. J., Bachman, L. F., Perkins, K., & Cohen, A. D. (1991). An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources. Language Testing, 8, 41–66. doi:10.1177/0265532 29100800104

Artemeva, N., & Fox, J. (2010). Awareness vs. production: Probing students’ antecedent genre knowledge. Journal of Written and Business Communication, 24, 476–515.

Ascención, Y. (2005). Validation of reading-to-write assessment tasks performed by second language learners (Unpublished doctoral dissertation). Northern Arizona University, Flagstaff.

Asencio´n, Y. (2004). Validation of reading-to-write assessment tasks performed by second language learners. Unpublished doctoral dissertation, Northern Arizona University, Flagstaff, AZ.

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

Bachman, L. F. (2002). Some reflections on task-based language performance assessment. Language Testing, 19, 453–476. doi: 10.1191/0265532202lt240oa

Bachman, L. F., & Palmer, A. S. (1996). Language testing in practice. Oxford, England: Oxford University Press.

Benson, P. & Voller, P. 1997. Autonomy and Independence in Language Learning. London: Longman.

Bereiter, C. (2002). Education and mind in the knowledge age. Mahwah, NJ: Lawrence Erlbaum. Byrnes, H. (2008). Assessing content and language. In E.

Bertelson, P. (1986). The onset of literacy: Liminal remarks. Cognition, 24, 1-30

Braine, G. (2001). When an exit test fails. System, 29, 221–234. doi:10.1016/ S0346251X(01)00009-4

Brown, H.D. (1991). Breaking the language barrier: Creating your own pathway to success. Yarmouth, ME: Intercultural Press.

Brown, J. D., Hilgers, T., & Marsella, J. (1991). Essay prompts and topics: Minimizing the effect of mean differences. Written Communication, 8, 533–56.

Byrne, B. M. (1994). Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming. Thousand Oaks, CA: Sage.

Byrnes, H., Maxim, H., & Norris, J. (2010). Realizing advanced foreign language writing development in collegiate education: Curricular design, pedagogy, assessment (Monograph). Modern Language Journal, 94, Suppl. 1.

Campbell, C. (1990). Writing with others’ words: Using background reading text in academic compositions. In B. Kroll (Ed.), Second language writing (pp. 211– 230). Cambridge, England: Cambridge University Press.

Carrasquillo, A.L. (1993) Whole native language instruction for limited-English-proficient students. In Angela Carrasquillo and Carolyn Hedley (Eds.), Whole Language and the Bilingual Learner (pp.3-19). Norwood, NJ:

Carroll, J. B. (1975). The teaching of French as a foreign language in eight countries. New York, NY: John Wiley and Sons.

Chalhoub-Deville, M. (2003). Second language interaction: Current perspectives and future trends. Language Testing, 20, 369–83.

Chalhoub-Deville, M., & Deville, C. (2005). Looking back at and forward to what language testers measure. In E. Hinkel (Ed.), Handbook of research on second language teaching and learning (pp. 815–32). Mahwah, NJ: Erlbaum.

Chamot, A. U. (1987). The learning strategies of ESL students. In A. L. Wenden, & J. Rubin (Eds.), Learner strategies in language learning (pp. 71–84). Englewood Cliffs, NJ: Prentice Hall.

Charge, N., & Taylor, L. B. (1997). Recent developments in IELTS. ELT Journal, 51, 374–380. doi:10.1093/elt/51.4.374

Cho, Y. (2003). Assessing writing: Are we bound by only one method? Assessing Writing, 8, 165–191. doi:10.1016/S1075-2935(03)00018-7

Cohen, A. D. (1998). Strategies and processes in test taking and SLA. In M. H. Long, & J. C. Richards (Eds.), Interfaces between second language acquisition and language testing research (pp. 90–111). Cambridge, England: Cambridge University Press.

Cohen, A. D. (2006). The coming of age of research on test-taking strategies. Language Assessment Quarterly, 3, 307–331. doi:10.1080/15434300701333129

Cohen, A. D., & Pinilla-Herrera, A. (2009). Communicating grammatically: Constructing a learner strategies website for Spanish. In T. Kao, & Y. Lin (Eds.), A new look at language teaching and testing: English as subject and vehicle (pp. 63–83). Taipei, Taiwan: Language Training and Testing Center.

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. Language Testing, 24, 209– 250. doi: 10.1177/0265532207076364

Colpin, M., & Gysen, S. (2006). Developing and introducing task-based language tests. In van den Branden, K. (2006), Task-based language education (pp. 151–74). New York, NY: Cambridge University Press.

Cope, B., & Kalantazis, M. (Eds.) (2000). Multiliteracies: Literacy learning and the design of social futures. London, England: Routledge.

Crocker, L. (2006). Preparing examinees for test taking: Guidelines for test developers. In S. M. Downing, & T. M. Haladyna (Eds.), Handbook of test development (pp. 115–130). Mahwah, NJ: Lawrence Erlbaum.

Cumming, A. (2013). Assessing integrated writing tasks for academic purposes: Promises and perils. Language Assessment Quarterly, 10, 1–18. Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. Language Testing, 21, 159–97. (See also TOEFL Monograph Report 26 at [http://www.ets.org/research/policy\_ research\_reports/rm-04-05\_toefl-ms-26](http://www.ets.org/research/policy_%20research_reports/rm-04-05_toefl-ms-26))

Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. Language Testing, 21, 159–197. doi:10.1191/0265532204lt278oa

Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Difference in written discourse in independent and integrated prototype tasks for next generation TOEFL. Assessing Writing, 10, 5–43. doi:10.1016/j.asw. 2005.02.001

Cumming, A., Rebuffot, J., & Ledwell, M. (1989). Reading and summarizing challenging texts in first and second languages. Reading and Writing: An Interdisciplinary Journal, 2, 201–19.

Currie, P. (1998). Staying out of trouble: Apparent plagiarism and academic survival. Journal of Second Language Writing, 7, 1–18. doi:10.1016/S1060-3743(98) 90003-0

Davies, A. (2008). Assessing academic English: Testing English proficiency 1950–2005, the IELTS solution. Cambridge, England: Cambridge University Press.

De Fina, A., Anstendig, L., & De Lawter, K. (1991). Alternative integrated reading/writing assessment and curriculum design. Journal of Reading, 34, 354–9.

Delaney, Y. A. (2008). Investigating the reading-to-write construct. Journal of English for Academic Purposes, 7, 140–50.

Do¨rnyei, Z. (2003). Questionnaires in second language research: Construction, administration and processing. Mahwah, NJ: Lawrence Erlbaum.

Dulay, H. , Burt, M., & Krashen, S. (1982). Language two. Oxford, UK: Oxford University Press. Farris, P. & Kaczmarski, D. (1988). Whole language, a closer look. Contemporary Education, 59(2), 77-81

Erlbaum. van Lier, L. (1989). Reeling, writhing, drawling, stretching, and fainting in coils: Oral interviews as conversation. TESOL Quarterly, 23, 489–508.

Esmaeili, H. (2002). Integrated reading and writing tasks and ESL students’ reading and writing performance in an English language test. Canadian Modern Language Review, 58, 599–622. doi:10.3138/cmlr.58.4.599

Feak, C., & Dobson, B. (1996). Building on the impromptu: A source-based academic writing assessment. College ESL, 6(1), 73–84.

Flowerdew, J., & Li, Y. (2007). Language re-use among Chinese apprentice scientists writing for publication. Applied Linguistics, 28, 440–65.

Fox, J. (2003). From products to process: An ecological approach to bias detection. International Journal of Testing, 3(1), 21–47. doi:10.1207/S15327574IJT 0301\_2

Fox, J. (2004). Test decisions over time: Tracking validity. Language Testing, 21, 437–465. doi:10.1191/0265532204lt292oa

Freeman, D. & Freeman, Y. 1992). Whole Language for Second Language Learners. Portsmouth, NH: Heinemann.

Frost, K., Elder, C., & Wigglesworth, G. (2012). Investigating the validity of an integrated listening–speaking task: A discourse-based analysis of test takers’ oral performances. Language Testing, 29, 345–69.

Gass, S.M., & Selinker, L. (2001). Second Language Acquisition: An Introductory Course. London, UK: Lawrence Erlbaum.

Gebril, A. (2009). Score generalizability of academic writing tasks: Does one test method fit it all? Language Testing, 26, 507–531. doi:10.1177/0265532209340188

Gebril, A., & Plakans, L. (2009). Investigating source use, discourse features, and process in integrated writing tests. Spaan Working Papers in Second or Foreign Language Assessment, 7, 47–84.

Gilham, B. (2000). Developing a questionnaire. London, England: Continuum.

Gough, P. & Tunmer, W. (1986). Decoding, reading, and reading disability. Remedial and Special Education, 7, 6-10.

Grabe, W., & Kaplan, R. B. (1996). Theory and practice of writing. New York, NY: Longman.

Green, M.L. (1989). 'Methods for teaching non-English speaking adults, Adult Learning, 1(3):25-27.

Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in highstakes testing. Educational Measurement: Issues and Practice, 23(1), 17–27. doi: 10.1111/j.1745-3992.2004.tb00149.x

Hamp-Lyons, L. (2001). Fourth generation writing assessment. In T. Silva, & P. K. Matsuda (Eds.), On second language writing (pp. 117–128). Mahwah, NJ: Lawrence Erlbaum.

Harley, B. (1986). Age in Second Language Acquisition. San Diego, CA: College- Hill.

Harley, B., Allen, P., Cummins, J., & Swain, M. (Eds.). (1990). The development of second language proficiency. New York, NY: Cambridge University Press.

Harwood, N., & Petric, B. (2012). Performance in the citing behavior of two student writers. Written Communication, 29, 55–103.

Hawkey, R. (2004). A modular approach to testing English language skills: The development of the Certificates in English Language Skills (CELS) examination. Cambridge, England: Cambridge University Press.

Hidi, S., & Anderson, V. (1986). Producing written summaries: Task demands, cognitive operations, and implications for instruction. Review of Educational Research, 56, 473– 93.

Higgins, D., Zechner, K., & Williamson, D. (2008). Automated scoring of spontaneous speech using SpeechRater V 1.0 (ETS Research Report 08-62). Princeton, NJ: Educational Testing Service.

Hillocks, G., Jr. (2002). The testing trap: How state assessments of writing control learning. New York, NY: Teachers College Press.

Hinkel, E. (2006). Current perspectives on teaching the four skills. TESOL Quarterly, 40, 109–131. doi:10.2307/40264513

Hirvela, A. (2004). Connecting reading and writing in second language writing instruction. Ann Arbor: University of Michigan Press.

Homburg, T. J., & Spaan, M. C. (1981). ESL reading proficiency assessment: Testing strategies. In M. Hines, & W. Rutherford (Eds.), On TESOL ‘81 (pp. 25–33). Washington, DC: TESOL.

Horwitz, E. K. (2008). Becoming a second language teacher: A practical guide to second language learning and teaching. Boston, MA: Pearson.

Howard, R. M. (1995). Plagiarisms, authorships, and the academic death penalty. College English, 57, 788–806. doi:10.2307/378403

Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6, 1–55. doi:10.1080/10705519909540118

Huang, H. T. (2010). Modeling the relationships among topical knowledge, anxiety, and integrated speaking test performance: A structural equation modeling approach (Unpublished doctoral dissertation). University of Texas at Austin.

Hyland, T. A. (2005). A study of the practices of EL1 and EL2 students in reporting information from sources in a timed writing proficiency assessment. Unpublished doctoral dissertation, University of Toronto.

Ibrahim, M. H. (1993). Integrating grammatical structures with communicative practice through situationally based dialogues. Occasional Papers in the Development of English Language Education, 17, 81-106.

Jo¨reskog, K. G., & So¨rbom, D. (1989). LISREL 7: A guide to the program and applications (2nd ed.). Chicago, IL: SPSS.

Kenny, D. A. (1979). Correlation and causality. New York, NY: John Wiley & Sons.

Kim, J. O., & Mueller, C. W. (1978). Introduction to factor analysis: What it is and how to do it. Newbury Park, CA: Sage.

Kintsch, W. (1998). Comprehension: A paradigm for cognition. New York, NY: Cambridge University Press.

Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. TESOL Quarterly, 25, 105–121. doi:10.2307/3587030

Knoch, U. (2009). Diagnostic writing assessment: The development and validation of a rating scale. Frankfurt, Germany: Peter Lang.

Koda, K. (2005). Insights into second language reading. Cambridge, England: Cambridge University Press.

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

Krashen, S.D. (1981). Second Language Acquisition and Second Language Learning. London, UK: Pergamon.

Krashen, S.D. (1987). Applications of psycho-linguistic research to the classroom. In M.H.

Kunnan, A. J. (1998). An introduction to structural equation modeling for language assessment research. Language Testing, 15, 295–332. doi:10.1177/ 026553229801500302

Lado, R. (1961). Language testing: The construction and use of foreign language tests. London, England: Longman.

Lee, Y.-W., & Kantor, R. (2005). Dependability of new ESL writing test scores: Evaluating prototype tasks and alternative rating schemes (Report No. MS-31). Princeton, NJ: Educational Testing Service.

Leki, I. (2007). Undergraduates in a second language: Challenges and complexities of academic literacy development. New York, NY: Erlbaum.

Leki, I., & Carson, J. (1997). “Completely different worlds”: EAP and the writing experiences of ESL students in university courses. TESOL Quarterly, 31(1), 39–69.

Leki, I., Cumming, A., & Silva, T. (2008). A synthesis of research on second language writing in English. New York, NY: Routledge.

Lewkowicz, J. (1994). Writing from sources: Does source material help or hinder students’ performance? Retrieved from http://www.eric.ed.gov/PDFS/ ED386050.pdf

Lewkowicz, J. A. (1997). The integrated testing of a second language. In C. Clapham & D. Corson (Eds.), Encyclopedia of language and education (pp. 121–30). Dordrecht, Netherlands: Kluwer.

Lewkowicz, J. A. (2000). Authenticity in language testing: Some outstanding questions. Language Testing, 17, 43–64.

Long & J.C. Richards (Eds). Methodology in TESOL: A Book of Readings. Boston: Heinle & Heinle.

Long, M.H. (1983). Does Second Language Instruction make a Difference? A review of research, TESOL Quarterly, 17(3), 359-382.

Mateos, M., Villalon, R., de Dios, M. J., & Martin, E. (2007). Reading and writing tasks on different university degree courses: What do the students say they do? Studies in Higher Education, 32, 489–510. doi:10.1080/03075070701476183

McDaniel, M. & Pressley, M. (1986). Keyword and context instruction of new vocabulary meanings: Effects of text comprehension and memory. Journal of Educational Psychology, 81(2), 204-213.

Messick, S. (1989). Validity. In R. L. Linn (Ed.), Educational measurement (3rd ed., pp. 13–104). New York, NY: American Council on Education and Macmillan.

Messick, S. (1994). The interplay of evidence and consequences in the validation of performance assessments. Educational Researcher, 23(2), 13–23. doi:10.3102/ 0013189X023002013

Mislevy, R., & Yin, C. (2009). If language is a complex adaptive system, what is language assessment? In N. Ellis & D. Larsen-Freeman (Eds.), Language as a complex adaptive system. Supplement to Language Learning, 59, 249–67.

Morrow, K. (1977). Techniques of evaluation for a notional syllabus. London, England: Royal Society of Arts.

Nation, P. (1989). 'Speaking activities: Five features', ELT Journal, 43(1): 24-29

Norris, J. (2002). Interpretations, intended uses and designs in task-based language assessment. Language Testing, 19, 337–46.

Nunan, D. (1989). Designing Tasks for the Communicative Classroom. Cambridge: Cambridge University Press Omaggio, A. 1986. 'Successful language learners: What do we know about them?', ERIC / CLL News Bulletin, May, 2-3 Oxford, R. 2001.

O’Malley, J. M., & Chamot, A. U. (1990). Learning strategies in second language acquisition. Cambridge, England: Cambridge University Press.

Oller, J. W. (1979). Language tests at school. Harlow, England: Longman.

Oller, J. W. (1983). Evidence for a general language profi ciency factor: An expectancy grammar. In J. W. Oller (Ed.), Issues in language testing research (pp. 3–10). Rowley, MA: Newbury House.

Oxford, R. L. (2001). Integrated skills in the ESL/EFL classroom. ESL Magazine, 4 (1), 18–20.

Peirce, B. (1992). Demystifying the TOEFL reading test. TESOL Quarterly, 26, 665– 89.

Pennycook, A. (1996). Borrowing others’ words: Texts, ownership, memory, and plagiarism. TESOL Quarterly, 30, 210–230. doi:10.2307/3588141

Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. Language Testing, 20, 26–56. doi:10.1191/0265532203lt243oa

Plakans, L. (2008). Comparing composing processes in writing-only and readingto-write test tasks. Assessing Writing, 13, 111–129. doi:10.1016/j.asw.2008.07.001

Plakans, L. (2009a). Discourse synthesis in integrated second language writing assessment. Language Testing, 26, 561–587. doi:10.1177/0265532209340192

Plakans, L. (2009b). The role of reading strategies in integrated L2 writing tasks. Journal of English for Academic Purposes, 8, 252–266. doi:10.1016/j.jeap. 2009.05.001

Plakans, L. (2010). Independent vs. integrated tasks: A comparison of task representation. TESOL Quarterly, 44(1), 185–94.

Plakans, L., & Gebril, A. (2012). A close investigation into source use in integrated second language writing tasks. Assessing Writing, 17, 18–34.

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 modeling approach. Language Testing, 15, 333–379. doi:10.1177/0265532298 01500303

Raimes, A. (1990). The TOEFL test of written English: Causes for concern. TESOL Quarterly, 24, 427–42.

Read, J. (1990). Providing relevant content in an EAP writing test. English for Specific Purposes, 9, 109–121. doi:10.1016/0889-4906(90)90002-T

Richards, J. (1983). Communicative needs in second and foreign language learning. English Language Teaching Journal (2).

Rosenfeld, M., Leung, S., & Oltman, P. (2001). The reading, writing, speaking, and listening tasks important for academic success at the undergraduate and graduate levels (TOEFL Monograph Report 21). Princeton, NJ: Educational Testing Service.

Rubin, J., Chamot, A. U., Harris, V., & Anderson, N. J. (2007). Intervening in the use of strategies. In A. D. Cohen, & E. Macaro (Eds.), Language learner strategies: 30 years of research and practice (pp. 141–160). Oxford, England: Oxford University Press.

Ruiz-Funes, M. (2001). Task representation in foreign language reading-to-write. Foreign Language Annals, 34(3), 226–34.

Sawaki, Y., Quinlin, T., & Lee, Y. (2013). Understanding learner strengths and weaknesses: Assessing performance on an integrated writing task. Language Assessment Quarterly, 10, 73–95.

Scarcella, R.C., & Oxford, R.L. (1992). The Tapestry of Language Learning: The Individual in the Communicative Classroom. Boston: Heinle & Heinle.

Shi, L. (2004). Textual borrowing in second-language writing. Written Communication, 21, 171–200. doi:10.1177/0741088303262846

Shi, L. (2010). Textual appropriation and citing behaviors of university undergraduates. Applied Linguistics, 31, 1–24.

Shohamy & N. H. Hornberger (Eds.), Encyclopedia of language and education. Volume 7: Language testing and assessment (2nd ed., pp. 37–52). New York, NY: Springer.

Spivey, N. (1984). Discourse synthesis: Constructing texts in reading and writing. Newark, DE: International Reading Association.

Spivey, N. (1990). Transforming texts: Constructive processes in reading and writing. Written Communication, 7, 256–287. doi:10.1177/0741088390007002004

Spivey, N. (1997). The constructivist metaphor: Reading, writing, and the making of meaning. New York, NY: Academic Press.

Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. Reading Research Quarterly, 24, 7–26.

Stein, V. (1990). Exploring the cognition of reading-to-write. In L. Flower, V. Stein, J. Ackerman, M. J. Kantz, K. McCormick, & W. C. Peck (Eds.), Reading-to-write: Exploring a cognitive and social process (pp. 119–143). New York, NY: Oxford University Press.

Sternglass, M. (1997). Time to know them: A longitudinal study of writing and learning at the college level. Mahwah, NJ:

Temple, C. And Gillet, J. W. (1984). Language Arts: Learning Processes and Teaching Practices. Boston: Little Brown and Company Ltd.

Trites, L., & McGroaty, M. (2005). Reading to learn and reading to integrate: New tasks for reading comprehension tests? Language Testing, 22(2), 174–210.

Upshur, J. A., & Turner, C. (1999). Systematic effects in the rating of second-language speaking ability: Test method and learner discourse. Language Testing, 16, 82–111. doi:10.1177/026553229901600105

Vacca, R., & Vacca, L. (2005). Content area reading: Literacy and learning across the curriculum (8th ed.). New York: Addison-Wesley.

Vygotsky, L.S. (1978). Mind in Society: The Development of higher psychological processes. Cambridge, MA: Harvard University Press.

Watanabe, Y. (2001). Read-to-write tasks for the assessment of second language academic writing skills: Investigating text features and rater reactions (Unpublished doctoral dissertation). University of Hawaii.

Weigle, S. C. (1999). Investigating rater/prompt interactions in writing assessment: Quantitative and qualitative approaches. Assessing Writing, 6, 145–178. doi:10.1016/S1075-2935(00)00010-6

Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. Assessing Writing, 9, 27–55. doi:10.1016/j. asw.2004.01.002

Wenden, A. L. (1991). Learner strategies for learner autonomy. Englewood Cliffs, NJ: Prentice Hall.

Wesche, B. (1987). Second language performance testing: The Ontario test of ESL as an example. Language Testing, 4, 28–47. doi:10.1177/026553228700400103

Wilhoit, D. (1994). Enhancing oral skills: A practical and systematic approach. English Language Teaching Forum, 32 (4), 32-36.

Yang, H. C. (2009). Exploring the complexity of second language writers’ strategy use and performance on an integrated writing test through SEM and qualitative approaches (Unpublished doctoral dissertation). University of Texas at Austin.

Yang, H., & Plakans, L. (2012). Second language writers’ strategy use and performance on an integrated reading-listening-writing task. TESOL Quarterly, 46, 80–103.

Yu, G. (2008). Reading to summarize in English and Chinese: A tale of two languages? Language Testing, 25, 521–51.

Yu, G. (2009). The shifting sands in the effects of source text summarizability on summary writing. Assessing Writing, 14, 116–37.

Yu, G. (2013). The use of summarization tasks: Some conceptual and lexical analyses. Language Assessment Quarterly, 10, 96–109. Xi, X.,