

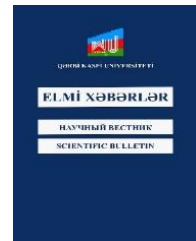
Scientific Bulletin, № 2, 2021, pages 48-58

Joseph Daubenmire

Western Caspian University

DOI: doi.org/10.54414/vrbs2410

e-ISSN: 2789-4614



A SELECTED REVIEW OF L2 ASSESSMENT RESEARCH LITERATURE IN SECOND LANGUAGE LEARNING

ABSTRACT

Assessment in second language learning is a vibrant field of research endeavor, and the subject matter itself is quite important. In this paper, fourteen geographically dispersed studies are presented which give the reader a sense of the breadth of the assessment research literature.

Keywords: Applied Linguistics; Second Language Assessment

İKİNCİ DİL ÖYRƏNƏN TƏLƏBƏLƏRİN BİLİKLƏRİNİYOXLAMA SAHƏSİNDƏ TƏDQIQAT

ОБЗОР ТЕСТИРОВАНИЯ СТУДЕНТОВ, ИЗУЧАЮЩИХ ВТОРОЙ ЯЗЫК

XÜLASƏ

РЕЗЮМЕ

İkinci dil tədrisində biliklərin yoxlanması çox həssas sahədir və bu mövzu çox vacibdir. İkinci dil öyrənənlərin ideyaları mövcüd olmayan formatda təqdim olunur. Məqsədə nail olmaq üçün tədqiqat jurnalı lazımdır. Burada tədqiqat ədəbiyyatının on dörd nümunəsi təqdim olunur. Onlar testlə bağlı tədqiqat ədəbiyyatının genişliyi haqqında anlayış yaradır.

Açar sözlər: dilçilik, ikinci dil, biliyinin yoxlanması

Тестирование в изучении второго языка достаточно деликатная область исследования, а сам предмет очень важен. Многие идеи и достижения в технике тестирования, изучающих второй язык, представлены в новом формате. Для достижения цели нужен подробный журнал исследования. Здесь представлены четырнадцать географически разбросанных учений, которые дают представление о широте исследовательской литературы тестирования.

Ключевые слова: Тестирование второго языка, прикладная лингвистик

Introduction: The following journal articles are selected from a composite of peer reviewed research conducted with learners from France, China, Germany, Iran, Turkey, Japan, Somalia, Eritrea, Portugal and Pakistan and which highlight some of the more interesting ideas that might serve as a basis for both L2 assessment practices and L2 research. The list is neither topical nor alphabetical but is arranged in such an order as to encourage flexibility on the part of English teachers and applied linguistics as they go about the development of their own assessment activities.

Relevance of Topic: The reviews are intended as remembrances which keep alive for the reader the notion that the science of L2 assessment is actually very broad and that creative approaches are the norm, rather than the exception. It is better to have such an overview at one's fingertips, in a journal, than filed away somewhere in class field notes.

Literature Review:

1. Collings, L. & White, J. (2011). An Intensive Look at Intensity and Language Learning. TESOL Quarterly, 45, 1, 106-133.

This longitudinal study is interesting because of the procedures or the outcome, as much as it is for the creative variety of testing instruments that were adapted in order to carry out the study. This approach will be useful in a situation where there is a need for suggestions on how to develop testing instruments to *measure specific skills*. The study investigated whether different distributions of instructional time would have differential effects on the acquisition of English by young (aged 11–12 years) French-speaking learners. Overall, the findings showed substantial progress over time for both groups, with no clear learning advantage for either concentrating or distributing the intensive experience. These results are consistent with other research comparing the effects of massed and distributed conditions on the learning of complex skills in other domains.

Pretests: The three pretests were an aural vocabulary recognition (AVR) test, a dictation test, and a cloze test. The AVR required students to match 80 words spoken on a tape to pictures on a series of pages in a test booklet. Scores were based on the total number of words correctly identified. The dictation was a 50-word text about a vacation.

Longitudinal Tests: Vocabulary knowledge. A yes–no vocabulary recognition test evaluated familiarity with the 1,000 most frequent words of English. It consisted of a checklist that contained 120 real words and 60 non-sense words.

Narrative writing. A picture-prompted written narrative task was used at each testing session (adapted from Collins et al, 1999; see also Lightbrown, Halter, White, & Horst, 2002). Four different pictures were chosen to match learners' developing vocabulary. The prompt in each case was 'imagine what is happening now, what happened before, and what is going to happen next'.

Learners were given 15-20 minutes to write. There were two measures for this task: fluency (number of words in the text) and grammatical knowledge (use of verb inflections).

Listening skills. The students' listening skills improve dramatically in these intensive programs. To measure communicative effectiveness, two timed, paired oral interaction measures were used with a stratified subset of 12 students from each class.

Conclusion: Findings also pointed to the importance in classroom-based research of including qualitative measures (observing classrooms, obtaining information from teachers) to document the instructional contexts, even when those contexts were familiar.

2. Yang, Q., Miller, M. & Bai, K. (2011). Testing the Improvement of English as a Foreign Language Instruction Among Chinese College Students Through Computerized Graphic Visuals. TESOL Quarterly, 2011, 45, 1, 170-182.

This study is informative both as a methodology of conducting research and in the way that the writers have structured their findings with practical application. It is an encouragement to teachers today to look for ways to include animated visual into their EFL teaching.

For nearly 40 years there has been agreement that computer generated graphics can impact learning by allowing the dual coding of material by the learner to strengthen and deepen memory, recall, and cognition. The process of verbal recall has been noted to be enhanced by pictorial exposure, allowing the connection and interplay between different parts of the brain, thus creating better memory and learning. Only with the emergence of computer technology has the visual element become more sophisticated, three-dimensional, and often, fluid and on-demand from the learner.

Methodology and Findings: The sample for the study included second-year college students in the departments of Chinese language and literature, political sciences, and history at a 4-year comprehensive university in China. One hundred and seventy (170) students were randomly selected. The findings of the current study supported that concrete visuals should be integrated with other visuals as a study aid to help students achieve a better learning outcome. These findings should help inform best practices for college faculty and should be shared and discussed among faculty as they prepare syllabi and lessons for their instruction. The findings should be considered by faculty in the United States and other countries as they work with students in their classes.

3. Hatipoğlu, Ç., & Algı, S. (2018). Catch a tiger by the toe: Modal hedges in EFL argumentative paragraphs. Educational Sciences: Theory & Practice, 18, 957–982. <http://dx.doi.org/10.12738/estp.2018.4.0373>

Writing argumentative paragraphs is challenging even in one's first language (L1) since in order to fulfil their goals writers need to carefully choose

among the available meta-discursive tools and skillfully balance their use. Writing in a foreign language (L2) is even more challenging because language learners are usually familiar only with a limited number of meta-discursive markers and functions. Therefore, when unsure, these novice L2 writers tend to fall back to old habits and transfer structures from L1 into their L2 texts. However, structures that are acceptable and may even be the norm in L1 may not be appropriate to use in L2.

The findings of this study show that modal hedges in English are a group of markers particularly problematic for second language learners as they are multifunctional, multilayered and culture dependent, and that some of the inappropriate uses or overuses of modals in L2 can stem from the employed teaching materials and/or lack of proper training related to this domain. The results emphasize once again how vital it is to find a place for the meta-discourse markers in the foreign language writing curricula as well as in the paragraph assessment rubrics used in the institutions.

4. Ozkan, Y. (2011). Assessment of Grammatical Competence Based on Authentic Texts. *International Journal of English Linguistics*, 1(2), 148-154.

Given the sharp contrasts between vocabulary and grammar features in grammar materials and the language used in the real world, this study employed authentic materials in grammar just as it appears in real life, using news text, with a group of 125 ELL's at the University of Cukurova, Turkey. The relevant portions of the grammatical features being taught were presented in underlined format and this allowed learners to easily identify, research, and reproduce learning in a testing format that was designed to reinforce learning, rather than simply assess what had been learned. As an alternative to the usual multiple choice testing that is predominant in grammar classes, the participants regarded this experience as a positive alternative in general, and for its positive washback effect on learning in particular. Nine themes emerge from using original text and authentic language.

5. Jafary, M., Nordin, N., & Mohajeri, R. (2012). The Effect of Dynamic versus Static Assessment on Syntactic Development of Iranian College Preparatory EFL Learners. *English Language Teaching*, 5(7), 149-157.

Two groups, experimental and control, age range was between 17 and 19, received instruction six hours a week, seventeen sessions. After determining the proficiency level of the participants, a forty-item test of grammar was administered as the pre-test to measure the participants' knowledge of grammar. A 30-item pre- and post-test was used. The main part of the study, i.e. treatment, made use of a mediation process which enables teachers to test-teach-test. An ANCOVA was run to compare experimental and control groups after the treatment period and to show whether post-test differences were due

to treatment effect (mediation) or any irrelevant variables. In order to answer the main question of the study: 'Is there any statistically significant difference between direct assessment and non-direct assessment procedures in L2 grammar instruction?'

There was a wide gap in their post-test scores, indicating the superiority of experimental group to control group in terms of their performance on a grammar test after the treatment period. The findings revealed that DA procedures had a significant and meaningful effect on the ease and feasibility of L2 grammar teaching and learning. The implication is that learners can benefit from DA-based mediation and that teacher intervention can be very instrumental in the process of L2 grammar instruction.

6. Marefat, H., & Shirazi, M. (2003). The impact of teaching direct learning strategies on the retention of vocabulary by EFL learners. Reading Matrix: An International Online Journal, 3(2), 47-62.

Teaching requires evaluation and teaching strategies are no exception. In this study, evaluation was made through a test constructed for this purpose. The test was developed with regard to the materials prepared for the treatment phase. The test included 89 items examining vocabulary knowledge of the learners. These vocabulary items were taught in the treatment phase by adopting cognitive, memory, and compensation strategies. The researchers examined impact of teaching learning strategies related to memory and cognition and their related functions on vocabulary retention, both short and long term.

Participants of the study were sixty female English Language Learners in upper teens. A pre experimental questionnaire was given to the participants to find if they already used such techniques before receiving any instruction. This also increased consciousness of their use. After the treatment, the participants took two equivalent tests with an interval of two weeks to find out the difference between their short term and long term retention of vocabulary. The results indicated that learners' strategy use in short-term retention far outweighs that in long-term retention. The results also portrayed the superiority of memory strategy use both in short and long term retention. The next most frequently used strategies were cognitive and compensation strategies respectively.

7. Marti, L. (2006). Indirectness and politeness in Turkish-German bilingual and Turkish monolingual requests. Journal of Pragmatics, 38(11), 1836-1869.

This study explored the following questions: Are Turkish-German bilingual returnees more direct than Turkish monolinguals? and How direct are Turkish speakers compared to speakers of other cultures or languages? The study conducted in Turkey is based on Turkish-German bilingual returnees, Turkish natives who have returned to Turkey from Germany after living there for varying periods of time. The purpose of the study is to see if pragmatic

performance has been influenced by pragmatic transference from German. A Discourse Completion Test is administered as an initial step to monolingual Turks and Turkish-German returnees.

To measure indirectness, a politeness ranking questionnaire was administered and it was found that indirectness and politeness were related, but not linearly linked. While the returnees were perceived as being more direct, it was likely due to different norms of social behavior between Turks and Germans. Good self-critique of methods and findings are included in this article. The author argues for the necessity to consider not only 'the said', but also 'the unsaid' responses of the informants, namely: deliberate choices of opting out; alternative solutions provided by informants; and attempts for negotiation while making a request. Reference is made to Goffman's Face Threatening Act avoidance strategies as attempts to mitigate directness.

Cross-cultural comparison appeared to be a prerequisite for investigations on a possible pragmatic transfer from German into the Turkish of returnees. One large study, the Cross Cultural Speech Act Realization Project (CCSARP), has been administered to university students in eight cultures or languages (later this number rose to 13) to investigate cross-cultural and intralingual variation in two speech acts: requests and apologies. Generally, Turkish speakers adopt quite direct strategies. This is also the case when Turkish speakers are compared to German speakers, who usually prefer indirect strategies.

Turkish is an agglutinative language that tends to a more direct level of requests and here the author references the Polish linguist Wierzbicka when applying theory to her findings: The English language reflects a characteristically Anglo-Saxon culture where there is special emphasis on the rights and autonomy of every individual, tolerance of individual idiosyncrasies and peculiarities and is described as a culture or tradition which respects privacy, approves of compromises and disapproves of dogmatism. A warning is issued not to rely only on theories that are based on influential languages, such as English. Even so, the results provide insight into the native Turkish and Turkish-German bilingual speakers' tendency to use direct or indirect requests in various situations

8. Gedik, N. (2012). The Optimum Blend: Affordances and Challenges of Blended Learning for Students. Turkish Online Journal of Qualitative Inquiry, July 2012, 3(3).

This study investigated students' perceptions of the most facilitative and the most challenging course features in a blended learning environment (BLE) identified as affordances and barriers so that focus could be directed toward eliminating or minimizing them. The term 'affordance' is used to identify those course features which were perceived as helpful to the students and is a sub

factor within the study of pragmatics. 'Barrier' is used for the challenging course features.

This study looks at the university as a place where communication takes place in a blended learning environment and it carries out this research to determine what the perceptions are with regard to affordances that are provided in that environment. The research design is a qualitative phenomenological study that is intended to yield explanations of lived experiences, rather than focusing on learning outcomes because understanding phenomena is much more interrelated with experiences. Data were collected from ten undergraduate students.

The findings indicate that for such blended courses, the instructor must be careful to sustain a balance in terms of students' workload and time devotion, support mechanisms and guidance, and assessment.

9. Nezakatgoo, B. (2011). Portfolio as a Viable Alternative in Writing Assessment. *Journal of Language Teaching & Research*, 2(4), 747-756.

This quantitative study looks at effectiveness of the portfolio to measure development of EFL students' progress with writing mechanics such as spelling, punctuation, capitalization, abbreviations, and numbering which are often dealt with in the revision or editing stages and this makes portfolio assessment aptly suited for writing assessment since students are required to write but can choose their own topic, audience, responses in the class, and revision strategies.

The results of the study confirmed that students whose work was evaluated by a portfolio system (portfolio-based assessment) had a significant reduction in their errors in mechanics of writing when compared to those students whose work was evaluated by the more traditional evaluation system (non-portfolio-based assessment).

10. AssiSuzuki, M. (2009). The Compatibility of L2 Learners' Assessment of Self-and Peer Revisions of Writing with Teachers' Assessment. *TESOL Quarterly*, Vol. 43, No. 1 (March 2009), pp. 137-148.

This study examines L2 learners' self-assessment of text changes on written drafts in self revisions and peer revisions, focusing on linguistic features of their repairs. Twenty-four Japanese university students completed TOEFL ITP which became the pre-test baseline. Participants were divided into two similar groups. Students assessed their text changes during self-revisions and text changes made with their partners during their revisions differently. The study has pedagogical significance for organizing self-assessment in the process-oriented writing classroom. The findings can be helpful for more effective instruction for self-assessment of L2 writing.

11. Han, C., Starkey H, & Green, A., (2010). The politics of ESOL (English for speakers of other languages): Implications for citizenship

and social justice. *International Journal of Lifelong Education*, Vol. 29, Iss. 1, 2010.

A number of countries in Europe use language tests to grant citizenship. UK immigrants with insufficient language skills participate in an ESOL and citizenship course to develop skills necessary for social integration. This policy is argued against because of the gatekeeping function and related social injustice as well as assimilation versus multicultural issues. A question remains as to how far citizenship education helped students to become citizens. This paper presents an interview and focus group study of a particular ESOL/citizenship course with students from Somalia, Eritrea, Turkey and Portugal who were five months into a one-year ESOL course. Inconsistencies and contradictions in policies and discourses around language as well as around changing funding were identified, particularly because these policies determined who could participate.

12. Jilani, R. (2009). *Problematizing High School Certificate Exam in Pakistan: A Washback Perspective. Reading Matrix: An International Online Journal*, 9(2), 175-183.

In Pakistan the high school certificate exam (HSC) has been used for 30 years and the exam has widespread educational and social implications. This paper exams those dynamics and reflects on the social washback impact and the need for serious attention in view of these implications. In Pakistani society, English has always been the language of power and a virtual class marker to identify an individual's social and educational status. Proficiency in English is arguably the most fundamental advantage for Pakistani youth. The test requires three hours of writing and this has given rise to a culture of rote learning since students only need to memorize and reproduce answers to questions frequently asked over the years. Acquiring of high scores in English is easy since gaining L2 proficiency is not necessary. Classroom teaching focuses on discrete-point grammar items and the translation of reading passages. This system allows little or no learning of English to actually take place in the classrooms.

13. Wiley, T. & Lukes, M. (1996). *English-Only and Standard English Ideologies in the U.S. TESOL Quarterly*, 30: 511–535.

This highlights English monolingualism vs. standard language ideology that position speakers of different varieties of the same language within a social hierarchy. Linguistic ideologies are not autonomous. They are linked to other social ideologies, the most prevalent of which is social mobility through education, and the instrumental role schools play in language assessment and classification schemes. These facilitate access to social capital for those who have mastered the standard language. Tracking students according to language ability is an essential component of social mobility. Schools stratify students based on their ability to use the standard by assigning those who speak Eng-

lish as an L2 or nondominant varieties of English to “remedial” educational tracks. The authors note that the real issue is not language, literacy, or education but power and a fear of heterogeneity, be it through language, behavior, or values and they exhort language professionals to contest policies and practices that perpetuate social inequities.

14. Shabani, K. (2016). Implications of Vygotsky’s sociocultural theory for second language (L2) assessment. *Cogent Education*, 3(1), 1242459. <https://doi.org/10.1080/2331186X.2016.1242459>

Lantolf and Poehner, echoing Vygotsky, contend that learners’ cognitive abilities are not stable but always in the state of flux implying that they can change through mediation and as educators we must feel accountable for their development. It is, therefore, unethical to act as “gate-keepers” only. Despite the presence of a rich and robust research literature reflecting more than 40 years of professional works in psychology and general education, the direct assessment approach in the context of second language studies is still in its infancy. However, given its strong epistemological background, it can serve as a potential site to trigger innovative works on different aspects of L2 research and even teacher education. Another use of DA in the context of applied linguistics is the effect of peer-to-peer mediation on learners’ mastery of different components of L2. Peer mediation could provide opportunities for learners to engage in interactive dialogs in which they can notice one another’s weak points and cooperatively think of appropriate strategies to resolve their linguistic uncertainties.

Discussion and Results:

The above literature gathered from various points on the global compass is by no means exhaustive, nor is it representative but it is a fair sampling of some of the areas of concern that applied linguists benefit from by being exposed to these ideas. These items of research also serve to impress the reader that L2 assessment is not a simple, or cut and dried instructional tool to be used rotely without thought to the impact which that tool has on learners.

In addition to the challenging and often interesting classroom dilemmas presented in the above readings, we must also take note of the serious lifelong inequities that result from one’s poor performance on high stakes learning tests which, while impossible to remedy, still are an area of huge concern to the socially conscious professional.

In addition to the eclectic responses presented above several other major issues come to mind. First, it’s important to insure that pre and post testing is properly normed and also state approved. Such testing can pinpoint difficulties with specific language deficits that need to be addressed and can also help to determine to what degree activities need to be modified to make language learning objectives attainable. Otherwise the problems that arise from testing specifically for discrete responses overlooks perhaps considerable amounts of

language knowledge attainment that exists but is not observed because it is not tested for. Such an approach of relying on one-shot assessment can greatly hamper student motivation. Under such circumstances, testing almost becomes counterproductive, given the important role of personal motivation related to learning.

As far as testing protocol is concerned, it's important that students be given appropriate "break in" time before being tested to reduce feelings of discomfort and shame. In cases where students seem to be struggling excessively, assessment in their L1 is also essential in order to rule out any possible learning disability that may exist. Going forward, it may be good to implement more research-based assessment techniques.

Regardless, there is merit in an increased measure of authentic language assessment based on increased authentic language production. This implies a process of relying on multiple measures over time, not once for all, which may work well in law school or medical school, but do not do justice in language learning assessment. Thus, such performance-based assessment can be built from conversations, portfolios, teacher notes, parental input and many other demonstrations of functional performance.

Assigning a grade to such a collection of work is a rational act that requires a creative rubric, but as a general rule, it's good when language assessment can be designed to measure and reward a student for what they do know, and not penalize them for what they don't know.

Resources:

Collings, L. & White, J. (2011). An Intensive Look at Intensity and Language Learning. *TESOL Quarterly*, 45, 1, 106-133

Gedik, N. (2012). The Optimum Blend: Affordances and Challenges of Blended Learning for Students. *Turkish Online Journal of Qualitative Inquiry*, July 2012, 3(3)

Han, C., Starkey H, & Green, A., (2010). The politics of ESOL (English for speakers of other languages): Implications for citizenship and social justice. *International Journal of Lifelong Education*, Vol. 29, Iss. 1, 2010

Hatipoğlu, Ç., & Algi, S. (2018). Catch a tiger by the toe: Modal hedges in EFL argumentative paragraphs. *Educational Sciences: Theory & Practice*, 18, 957–982. <http://dx.doi.org/10.12738/estp.2018.4.0373>

Jafary, M., Nordin, N., & Mohajeri, R. (2012). The Effect of Dynamic versus Static Assessment on Syntactic Development of Iranian College Preparatory EFL Learners. *English Language Teaching*, 5(7), 149-157

Jilani, R. (2009). Problematizing High School Certificate Exam in Pakistan: A Washback Perspective. *Reading Matrix: An International Online Journal*, 9(2), 175-183

Marefat, H., & Shirazi, M. (2003). The impact of teaching direct learning strategies on the retention of vocabulary by EFL learners. *Reading Matrix: An International Online Journal*, 3(2), 47-62

Marti, L. (2006). Indirectness and politeness in Turkish-German bilingual and Turkish monolingual requests. *Journal of Pragmatics*, 38(11), 1836-1869

Nezakatgoo, B. (2011). Portfolio as a Viable Alternative in Writing Assessment. *Journal of Language Teaching & Research*, 2(4), 747-756

Ozkan, Y. (2011). Assessment of Grammatical Competence Based on Authentic Texts. *International Journal of English Linguistics*, 1(2), 148-154

Shabani, K. (2016). Implications of Vygotsky's sociocultural theory for second language (L2) assessment. *Cogent Education*, 3(1), 1242459. <https://doi.org/10.1080/2331186X.2016>

Suzuki, M. (2009). The Compatibility of L2 Learners' Assessment of Self-and Peer Revisions of Writing with Teachers' Assessment. *TESOL Quarterly*, Vol. 43, No. 1 (March 2009), pp. 137-148

Wiley, T. & Lukes, M. (1996). English-Only and Standard English Ideologies in the U.S. *TESOL Quarterly*, 30: 511–535

Yang, Q., Miller, M. & Bai, K. (2011). Testing the Improvement of English as a Foreign Language Instruction Among Chinese College Students Through Computerized Graphic Visuals. *TESOL Quarterly*, 2011, 45, 1, 170-182