

Three decades of research in language teaching and testing: A thematic categorization

Esmaeeli, Fatemeh ✉

Shiraz University, Iran (esmali9531513@shirazu.ac.ir)

Received: 6 December 2020

Available Online: 27 January 2021

Revised: 18 January 2021

DOI: 10.5861/ijrse.2021.5080

Accepted: 23 January 2021

ISSN: 2243-7703

Online ISSN: 2243-7711

OPEN ACCESS



Abstract

Following a qualitative research design, the researcher tried to determine the research's focus in the second language teaching and learning as well as in the language testing during the last three decades. In effect, different categories covered in the two fields of language teaching and assessment were recognized and presented. It was concluded that the two fields have had a similar trend in the last three decades in which different categories have been covered in different portion in the 1990's and 2000's but the categories were covered in a relatively balanced portion in the 2010's. Apparently, the two fields are closely interrelated which means that each one is signified and extended in terms of theoretical, practical and foundational advances of the other one. In addition, exploring the contributions of Iranian researchers into the two fields of language teaching and assessment during the last three decades resulted in too insignificant contribution.

Keywords: qualitative design; language testing; language teaching; categorization; interrelatedness

Three decades of research in language teaching and testing: A thematic categorization

1. Introduction

Language as a symbolic cognitive structure is multidimensional which may be considered from the two dimensions of teaching and testing. In effect, the two dimensions are two interrelated dimensions of the same entity wherein each is signified in terms of theoretical and practical definitions and foundation of the other. Likewise, the close relationships between the two dimensions are paramount for instructional setting and, even teaching English as a second or foreign language has added to their interrelationships. In this regard, Brown's (2004) emphasizes that teaching imbricates assessment in evaluating learners' competence. Following the ongoing processes, the two dimensions of instruction and assessment are presented in order to evaluate the way language is acquired and produced. Accordingly, language assessment and teaching are interrelated. Hence, teaching is the outer circle involves the both assessment and test.

Richards (1990, p. 15), in a similar vein, considers testing as a main component in curriculum development especially in terms of needs' analysis and evaluation which has resulted in "the design and delivery of instruction as well as for the administration of the program itself". He adds that decision making "in the course of a language program's designing" and implementation is at least conditioned by the "use of tests" (Richards, 1990, p. 15). The unique characteristics of the field of teaching and assessment of EFL/ESL resulted in a multidisciplinary approach wherein different factors are interacting and interplaying to develop such a complex phenomenon, i.e. language. Considering the increasing complexities and interrelationships between and among the two dimensions of teaching and assessment are even apparent through considering the changes in theoretical perspectives regarding language.

In the field of second language acquisition, three schools of thought are paramount, namely, Structuralism and Behaviorism in 1940s and 1950s; Rationalism and Cognitive Psychology in 1960s and 1970s; and Constructivism in 1980s, 1990s and 2000s. There is a similar trend in the field of language testing with three theories of Structuralist approach in 1940s and 1950s, Integrative approach in 1970s and 1980s, and Communicative approach in 1980s, 1990s and 2000s. Considering the emerged theories in the two fields of language acquisition and language testing clarifies the close and dialectical relationship between the two fields in which any change in one influences the other. Recent attention to sociocultural theory as an offshoot of constructivism in the field of language education has resulted in considering dynamic assessment in the field of language testing. However, this issue may remind backwash effect which deals with the influence of testing on teaching; the observed relation between the two fields is more than backwash effect which is the undeniable nature of social science as an interdisciplinary filed.

Considering the effect of testing on teaching and learning, Shohamy Donitsa-Schmidt and Ferman (1996) argue about four notions of washback effect, measurement driven instruction, curriculum alignment and systemic validity with their focus on the relation between language acquisition and testing. Likewise, the trend in theories development in the two fields of language acquisition and testing as well as in different language skills denotes the fact that the two fields may be related dialectically. In this regard, Vey (2005, p. 1) argues that "assessment does not work in a vacuum. Educational values should not only drive why we assess outcomes and accountability but also be a process of improving students' goals in conjunction with common and agreed national learning goals". It means that assessment processes and learning processes are related. Likewise, assessment practices should reflect educational insights which are evolving and improving in line with learning practices. Such a perspective regarding testing and assessment sheds light on the purposefulness and monitoring role of assessment and testing which is a sine qua non of teaching.

Unfortunately, any study directly related to the present study was not found; however, several studies were

found which are worth mentioning as follows. Attributing the assessment-related studies into three fields of linguistics, psychometrics, and edumetrics, Farhady (2005) argues that each field gets its own feedback from the results of intra-disciplinary areas. He also shows that language-related models rooted in intra-disciplinary perspectives suffer from inadequacies due to their ignorance of an inter-disciplinary perspective. Suggesting edumetrics perspective, Farhady (2005) points out that the three dimensions of intra-disciplinary, inter-disciplinary and supra-disciplinary should be considered in research for any language ability model.

Arguing about the changes in models for studies and development, Bachman (2000, p. 1) “reviews developments in language testing research and practice over the past twenty years, and suggests some future directions in the areas of professionalizing the field and validation research”. Considering the aforementioned issues and by taking into account that the two fields of language teaching and assessment are interrelated and go hand in hand, it appears that changes in one field may happen in the other as well.

1.1 Objectives of the study

Considering a close and dialectical relationship between the two fields of language teaching and assessment in which it is assumed that both fields are reflecting language learning processes, it may be interesting to explore the trends observed in different fields during three decades in the top journals. In addition, it may be beneficial to investigate the contribution of Iranian researchers in the two fields.

1.2 Research questions

In meeting the objectives of the study several questions were raised.

- How has the research’s focus in the second language teaching and learning changed during the last three decades?
- How has the research’s focus in the language testing changed during the last three decades?
- How are the contributions of Iranian researchers into the two fields of language teaching and assessment during the last three decades?

1.3 Significance of the study

The results of the study may be of interest to all who are concerned with the two fields of language teaching and assessment. In effect, the study may clarify the changing trends in studies in the two fields during the last three decades giving consciousness and information to the teachers, learners, researchers or even syllabus and curriculum designers who are dealing with language in one way or another. In fact, the results of the study may reveal some issues and concerns covered in the journals during the last three decades which shed light on the changing trends especially in terms of assessment-driven approaches beneficial for educational settings. It may also provide us a classification regarding the dimensions considered and changed during the last three decades in the two fields. Moreover, the studies conducted in each field along with their trend and changes may be regarded as a kind of a feedback for another field which clarifies some dimension beneficial for educational purposes to be applied for improvement or even for research-based purposes.

2. Methodology

In this section, the methodology-related issues are clarified.

2.1 Design of the study

This study followed a qualitative research design which was conducted based on the categorization found in the data along with percentage in clarifying the trend. Hence, we did not have any priori variables or categories.

Likewise, some sort of numbers along with descriptive statistics was generated to facilitate drawing the conclusions. In effect, after selecting the journals, all the papers published during the last three decades in those journals were explored in terms of the field and dimension based on which the study was conducted during different decades and were labeled based on its focus and dimension. Accordingly, the labels and categories were extracted out of the explored published papers as our data. In another words, first the journals were selected and divided into two groups of teaching and testing. All the papers published during the last three decades were explored and coded in terms of topic, dimension, and theme. After that, the categories were designed in which the number of papers related to each category in each decade were numerated and tabulated. The extracted data were put into SPSS 22 to be explored in terms of descriptive statistics which were also presented using appropriate diagrams.

2.2 Materials

The source of data were the first ten top journals in the field of education—language teaching and assessment during the last three decades. Hence, the rationale for selecting the journals was the journals' reputation among the academic members especially through the judgment of five experts as well as through the ISI ranking provided by the SCImago Journal and Country Rank which is a research group from the Consejo Superior de Investigaciones Científicas (CSIC), University of Granada, Extremadura, Carlos III (Madrid) and Alcalá de Henares, and is dedicated to information analysis, representation and retrieval by means of visualization techniques. The included journals were from the two fields of language teaching, i.e. *Language learning*, *Applied Linguistics*, *TESOL Quarterly*, *Modern Language Journal*, *Language Teaching Research* and *ELT journal* and testing journals, i.e. *Language Testing*, *Assessing Writing*, *Papers in Language Testing* and *Assessment* and *Language Assessment Quarterly*.

2.3 Sampling method

For the purpose of the study the first ten journals were selected—six journals for language teaching (*Language learning*, *Applied Linguistics*, *TESOL Quarterly*, *Modern Language Journal*, *Language Teaching Research* and *ELT journal*) and four journals for language assessment (*Language Testing*, *Assessing Writing*, *Papers in Language Testing* and *Assessment* and *Language Assessment Quarterly*) through purposive sampling. Then all the selected journals were explored in terms of all the volumes and issues during the last three decades. The following table shows the selected journals along with the number of volumes and issues:

Table 1

The included Journals in terms of number of volumes, issues and papers

Fields	Journals	Vol.	Issue	Papers	Years	Total
Language Teaching	Language learning	28	131	707	1990-2017	4480 (79%)
	Applied Linguistics	28	120	928	1990-2017	
	TESOL Quarterly	28	104	826	1990-2017	
	Modern Language Journal	28	172	708	1990-2017	
	ELT journal	28	112	931	1990-2017	
	Language Teaching Research	21	81	380	1997-2017	
Language assessment	Language Testing	28	99	708	1990-2017	1206 (21%)
	Assessing Writing	33	65	243	1994-2017	
	Language Assessment Quarterly	14	54	212	2004-2017	
	Papers in Language Testing & Assessment	6	10	43	2012-2017	
Total	8	242	948	5686	1990-2017	5686

2.4 Data collection procedures and analysis

After selecting the journals, the papers published during the past three decades were explored in two fields of language teaching and assessment in order to categorize and classify the published papers under more limited

categories. The gathered categories were supplemented with the percentage under each category in order to determine the trend and its changes if any. Finally, the contribution of Iranian researchers was explored in each filed. In order to provide more dependable results, descriptive statistics and Chi-square were applied in comparing different changes and found categories.

3. Results

This section presents the results achieved from the close investigation of all the papers published in the last three decades in two fields of language teaching and testing. As it was mentioned all the papers were explored, coded and themed which resulted into 16 categories for language testing and 13 categories for language teaching. Table 2 illustrates the categories extracted out of testing-related journals in terms of decades, frequency, percentage and the contribution of Iranian authors to this filed.

Table 2

Language testing and assessment in terms of the three decades

	Category	Decade			Total	Author	
		1	2	3		1	2
1	Feedback & error treatment (teacher or peers)	1 (1%)	7 (2%)	34 (6%)	42 (4%)	1025 (98%)	20 (6)
2	Rubric construction, rating scales, checklists, standards, & criteria	3 (2%)	2 (1%)	24 (4%)	29 (3%)		shared (2%)
3	Reliability, validity & accuracy, generalizability, authenticity, latent and construct-related issues	31 (19%)	51 (15%)	94 (17%)	176 (17%)		
4	task prompts and their effective roles, & assessment mode	5 (3%)	6 (2%)	6 (1%)	17 (2%)		
5	Washback & Portfolio	12 (7%)	9 (3%)	14 (3%)	35 (3%)		
6	Test taking strategies & processes (cognitive and meta-cognitive strategies; style of thinking and learning)	8 (5%)	14 (4%)	14 (3%)	36 (3%)		
7	Correlational and comparative studies	2 (1%)	4 (1%)	19 (4%)	25 (2%)		
8	Discourse related issues (voice, its features, genres, linguistic features, identity, critical approaches, wider historical or political issues, etc.)	5 (3%)	21 (6%)	25 (5%)	51 (5%)		
9	Variables related to raters and rating and decision-making processes	10 (6%)	15 (4%)	58 (11%)	83 (8%)		
10	Variables related to format and scoring procedures and method as well as tools (test method, design, characteristics, measure and integrated writing tasks, Reflexivity, Automated related issues	37 (22%)	71 (21%)	118 (22%)	226 (22%)		
11	Theoretical research (review, critiques, etc.)	14 (8%)	85 (25%)	55 (10%)	154 (15%)		
12	Contextual issues (contextual situations, developmental studies, affective-related variables, dynamics, background)	5 (3%)	21 (6%)	29 (5%)	55 (5%)		
13	Attitudinal studies (teachers or learners)	3 (2%)	3 (1%)	11 (2%)	17 (2%)		
14	Language teacher assessment literacy (teacher assessment, professional standards, teacher-assessors, etc.)	0 (0%)	11 (%)	20 (4%)	31 (3%)		
15	Ethical consideration	5 (3%)	10 (3%)	1 (0%)	16 (2%)		
16	Language assessment in relation to educational contexts and teaching approaches as well as classroom research with pedagogical implications	26 (16%)	9 (3%)	17 (3%)	52 (5%)		
	Total	167 (16%)	339 (32%)	539 (52%)	1045 (100%)		

As the table shows the last three decades have similar trends in terms of different categories; however, some differences are observed. The most prominent categories in 1990's are, in effect, related to the reliability, validity, accuracy, generalizability, authenticity, latent and construct-related issues as well as to language assessment in relation to educational contexts and teaching approaches and classroom research with pedagogical implications. In 2000's, the most prominent category is related to the theoretical research wherein a domain of language testing is clarified, reviewed, criticized or a new theory is found. In 2010's, on the other hand, Feedback and error treatment, Rubric construction, rating scales, checklists, standards, and criteria as well as variables related to raters and rating and decision-making processes have received the most attention among the researchers. Figure 1 demonstrates the trend in the three decades pictorially.

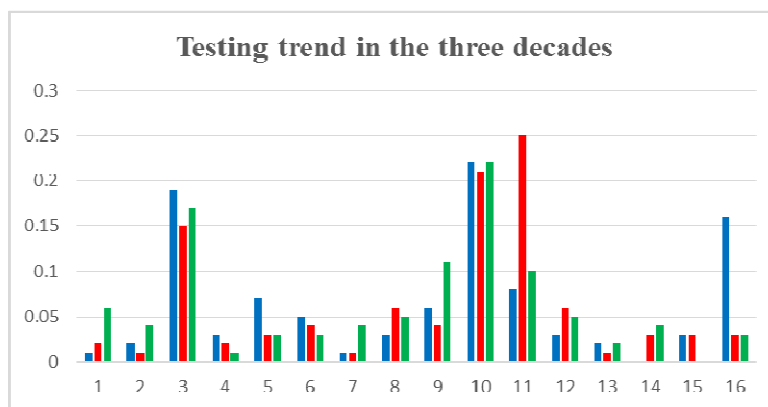


Figure 1. Language testing and assessment in terms of the three decades

Likewise, the contribution of Iranian authors is only 2% which 30% of this 2% contribution is related to Iranian authors studied and affiliated with abroad. Table 3, on the other hand, reveals the categories extracted from exploring the teaching-related journals.

Table 3

Language teaching in terms of the three decades

Category	Decade			Total	Author	
	1	2	3		1	2
1 Testing and evaluation	77 (5%)	95 (6%)	93 (4%)	265	4934 (99.99%)	43 (13 shared) (0.01%)
2 Professional Literacy (teacher's education & roles)	164 (11%)	231 (13%)	330 (15%)	725		
3 Curriculum design and development	84 (6%)	83 (5%)	80 (4%)	247		
4 Instructional methods, materials, and techniques	110 (7%)	126 (7%)	197 (9%)	433		
5 Social-related & Contextual issues (contextual situations, developmental studies, affective-related variables, dynamics, background, communication & collaboration-related research, pragmatics, language transfer, intercultural and inter-linguistic studies, interactional studies and usage-based research)	177 (12%)	218 (13%)	265 (12%)	660		
6 Skill-based studies and their components	101 (7%)	109 (6%)	132 (6%)	342		
7 Theoretical Studies (Review, critique, meta-analysis, etc.)	347 (23%)	305 (18%)	361 (16%)	1013		
8 Learning foreign and second languages (fluency, accuracy, meaning, grammar, pronunciation, etc.)	71 (5%)	74 (4%)	120 (5%)	265		
9 Language learning strategies, psycholinguistic-related issues & language learning processes (language processing, individual differences, cognitive, communicative, affective, compensative and meta-cognitive strategies; style of thinking and learning, working memory, metacognitive)	138 (9%)	175 (10%)	236 (11%)	549		
10 Discourse related issues (voice, its features, genres, linguistic features, identity, critical approaches, wider historical or political issues, etc.)	138 (9%)	190 (11%)	190 (9%)	518		
11 Task prompts and their effective roles, response roles, feedback, recast, correction, etc.	42 (3%)	73 (4%)	100 (5%)	215		
12 Attitudinal studies	18 (1%)	13 (1%)	22 (1%)	53		
13 Methodology based research	31 (2%)	32 (2%)	64 (3%)	127		
Total	1499 (28%)	1726 (32%)	2193 (40%)	5415		

Considering the extracted thirteen categories shows that the theoretical Studies which reviewed, clarified, criticized or presented a new theory or dimension is the most prominent theme in the 1990's. Moreover, during the 2000's, the three categories of testing and evaluation, social-related and Contextual as well as discourse related issues are the most prominent categories. Furthermore, the four categories of instructional methods, materials, and technique, task prompts and their effective roles, response roles, feedback, recast, correction, as well as methodology based research seem to be the most outstanding categories in 2010's.

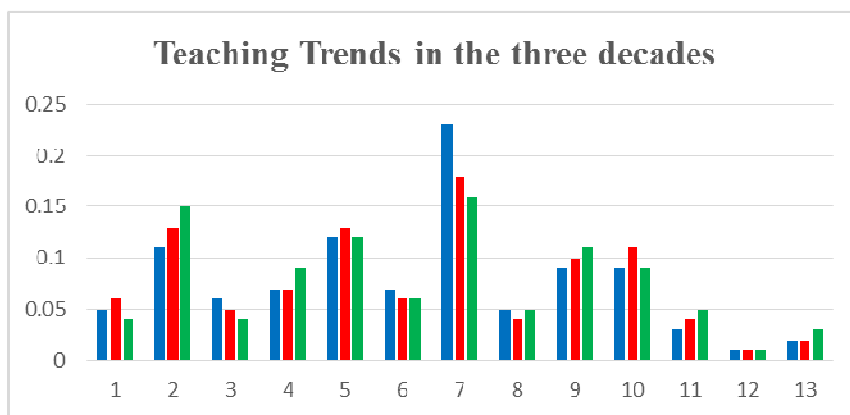


Figure 2. Language teaching in terms of the three decade

In addition, the contribution of Iranian author to the field of teaching is only about 0.01% of which 30% is related to the Iranian authors studied and affiliated with abroad. In a further step, a chi-square test was performed to examine the three decades in each field of testing and teaching as table 4 illustrates. The results of the chi-square test showed that there was no significant association between the decades and the field, $X^2(80) = 96$, $p = .107$.

Table 4

Chi-square test in terms of decades

	Chi-Square Tests	Value	df	Asymp. Sig. (2-sided)
Testing	Pearson Chi-Square	96.000	80	.107
	Likelihood Ratio	105.467	80	.030
	Linear-by-Linear Association	.000	1	1.000
Teaching	Pearson Chi-Square	39.000	48	.820
	Likelihood Ratio	49.648	48	.407
	Linear-by-Linear Association	.000	1	.998

Table 5 sheds light on the congruence between testing-related and teaching related categories which emphasizes on the fact each field is presenting some sort of feedback for another field which in its turn is following educational purposes especially for improvement in both fields and for provoking some research.

Table 5

Congruence between the categories of the two fields

Testing	Category	Teaching
➤ Feedback & error treatment;	➤	Task prompts and their effective roles, response roles, feedback, recast, correction, etc.
➤ Task prompts and their effective roles, & assessment mode;	➤	Curriculum design and development;
➤ Rubric construction, rating scales, checklists, standards, & criteria;	➤	Instructional methods, materials, and techniques;
➤ Washback & Portfolio;	➤	Language learning strategies, psycholinguistic-related issues & language learning processes;
➤ Test taking strategies & processes;	➤	Discourse related issues;
➤ Discourse related issues;	➤	Testing and evaluation;
➤ Variables related to raters and rating and decision-making processes	➤	Professional Literacy (teacher's education & roles);
➤ Variables related to format and scoring procedures and method as well as tools (test method, design, characteristics, measure and integrated writing tasks, Reflexivity, Automated related issues;	➤	Learning foreign and second languages (fluency, accuracy, meaning, grammar, pronunciation, etc.);
➤ Reliability, validity & accuracy, generalizability, authenticity, latent and construct-related issues;		
➤ Language teacher assessment literacy;		
➤ Language assessment in relation to educational contexts and teaching approaches as well as classroom research with pedagogical implications;		

Table 5 ...continued

		Category	
Testing		Teaching	
➤	Theoretical research;	➤	Theoretical Studies (Review, critique, meta-analysis, etc.);
➤	Contextual issues;	➤	Social-related & Contextual issues (contextual situations, developmental studies, affective-related variables, dynamics, background, communication & collaboration-related research, pragmatics, language transfer, intercultural and inter-linguistic studies, interactional studies and usage-based research);
➤	Attitudinal studies;	➤	Attitudinal studies;
➤	Ethical consideration;	➤	Methodology based research;

It is worth mentioning that different categories have been covered in different portion in the 1990's and 2000's but the categories are covered in a relatively balanced portion in the 2010's.

4. Discussion and conclusion

Having conducted the present study, 16 categories were recognized for the field of language testing in the past three decades. The recognized categories in the field of testing include:

- Feedback and error treatment
- Rubric construction, rating scales, checklists, standards, and criteria
- Reliability, validity and accuracy, generalizability, authenticity, latent and construct-related issues
- Task prompts and their effective roles, and assessment mode
- Washback and Portfolio
- Test taking strategies and processes
- Correlational and comparative studies
- Discourse related issues
- Variables related to raters and rating and decision-making processes
- Variables related to format and scoring procedures and method as well as tools
- Contextual issues
- Attitudinal studies
- Language teacher assessment literacy
- Ethical consideration
- Language assessment in relation to educational contexts and teaching approaches as well as classroom research with pedagogical implications

Likewise, 13 categories were recognized for the field of teaching which include:

- Testing and evaluation
- Professional Literacy
- Curriculum design and development
- Instructional methods, materials, and techniques
- Social-related and contextual issues
- Skill-based studies and their components
- Theoretical Studies
- Learning foreign and second languages
- Language learning strategies, psycholinguistic-related issues & language learning processes
- Discourse related issues
- Task prompts and their effective roles, response roles, feedback, recast, correction, etc.
- Attitudinal studies
- Methodology based research

Having considered the recognized categories in the two fields, it appears that there is a close and dialectical relationship between them. It denotes that both fields are reflecting language learning processes. Hence, a relatively similar trend is observed during the last three decades in the top journals in terms of the covered and published themes. Henceforth, the themes of feedback and error treatment as well as task prompts and their effective roles, and assessment mode in the field of language testing and assessment seem to be in line with the themes of task prompts and their effective roles, response roles, feedback, recast, correction, etc. in the field of

language teaching. Rubric construction, rating scales, checklists, standards, and criteria and washback and portfolio in the field of language testing are in line with the themes of curriculum design and development and instructional methods, materials, and techniques in the field of language teaching.

Likewise, exploring both fields revealed that both fields involve the themes of discourse related issues, theoretical studies, social and contextual issues as well as attitudinal theme. Another illustration is related to the fact that the two fields have considered the studies mainly related the opposite field. It means that some papers published in the testing-related journals are mainly related to the teaching issues and some papers published in the field of teaching are dealing with the variables mainly related to testing and assessment. It was concluded that the three decades are similar in terms of the research's focus in the second language teaching and learning as well as in terms of language testing and assessment during the last three decades. Apparently, the two fields are closely interrelated which means that each one is signified and extended in terms of theoretical, practical and foundational advances of the other one. Likewise, the contributions of Iranian researchers into the two fields of language teaching and assessment during the last three decades was revealed to be too marginal and almost zero especially for the field of language teaching.

5. References

- Bachman, L. F. (2000). Modern language testing at the turn of the century: Assuring that what we count counts. *Language testing*, 17(1), 1-42. <https://doi.org/10.1177/026553220001700101>
- Brown, H. D. (2004). *Language assessment: Principles and classroom practices*. Allyn & Bacon.
- Farhady, H. (2005). Language assessment: A linguametric perspective. *Language Assessment Quarterly: An International Journal*, 2(2), 147-164. https://doi.org/10.1207/s15434311laq0202_3
- Fulcher, G., & Davidson, F. (2007). *Language testing and assessment: An advanced resource book*. Routledge. <https://doi.org/10.4324/9780203449066>
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Kopriva, R. J. (2008). *Improving testing for English language learners*. Routledge.
- Osterlind, S. J. (2002). *Constructing test items: multiple-choice, constructed-response, performance, and other formats* (2nd ed.). Kluwer Academic Publishers.
- Shohamy, E., Donitsa-Schmidt, S., & Ferman, I. (1996). Test impact revisited: Washback effect over time. *Language Testing*, 13(3), 298-317. <https://doi.org/10.1177/026553229601300305>
- Vey, L. D. (2005). *Enhancing the relationship between learning and assessment*. University of Canberra. http://www.canberra.edu.au/researchrepository/file/edc0ca80-acb7-25b9-4fa8-6e03a01d4e27/1/introductory_pages.pdf

