

Bibliometric Analysis of the International Language Testing Research Through WOS Databases (2000-2020)

Yansong Feng^{1,*}, Jianghong Han¹

¹ School of Foreign Languages, Xuzhou University of Technology, Xuzhou, Jiangsu 221008, China

*Corresponding author. Email: fys720424@126.com

ABSTRACT

Using CiteSpace as the bibliometric analysis method, this paper analyzes the documents of three WOS databases (SSCI, CPCI-SSH and A&HCI) to systematically present the chronological development of the international language testing research during the past two decades. The visualized analyses display the knowledge map, research hotspots, development trajectory, research frontiers, important authors and references in language testing research. The paper provides a panorama of the new changes of language testing research in the past 20 years.

Keywords: CiteSpace, WOS, Language testing, Visualized analysis.

1. INTRODUCTION

It is generally believed that the publication of Language Testing by Lado in 1961 marked the official birth of international language testing research. Since then, language testing research has been developing quickly, and new theoretical viewpoints, research methods and research scopes have emerged and improved continuously. According to Davies (2014), only three articles have systematically reviewed the research development on language testing before 2000. Then, he analyzed the new developments of the international language testing research during the ten years from 2002 to 2012, and enumerated newly emerging research hotspots [7].

In China, Zhou Shanshan (2018) conducted a qualitative research of the 209 articles published in Language Assessment Quarterly and Language Testing between 2011 to 2015 and found that the four research hotspots during this period were the rater's rating performance, comprehensive testing, diagnostic evaluation, and classroom assessment. Language testing research is shifting from learning evaluation to learning promotion evaluation, and the social attributes of language testing are getting more and more attention [30]. Afterwards, Jin Yanhua (2020) once again reviewed the important

documents from 2008 to 2019 in Language Assessment Quarterly and Language Testing, and found seven hot research topics [12]. Literature retrieval shows that no scholar has conducted a comprehensive bibliometric analysis of the development trends, research hotspots and frontiers of international language testing in the past 20 years based on SSCI database. It is the purpose of this article.

2. DATA SOURCE AND MAIN RESEARCH TOOL

The data used in this article was derived from three WOS databases of SSCI, A&HCI and CPCI-SSH (2000-2020). The author used "language test*" or "language assess*" or "language evaluation" as the search subject, and "Article" or "Proceedings Paper" as the document type, and obtained a total of 2465 papers. Then, the publications were screened to exclude research papers in medical and clinical fields, and only literature related to language education was left. Eventually, a total of 529 articles from 42 kinds of publications were obtained for this paper.

In this paper, the author mainly used CiteSpace(5.7.R5W) as the data processing tool. By turning the tremendous literature data into

visualized language, CiteSpace can display the evolution process of a knowledge domain on a citation network map, and automatically identify the research frontiers represented by the citation node literature and co-citation clustering on the map as the knowledge base, showing the interpretability of the map itself [17]. This paper will present a bibliometric report on the development of language testing research in the past two decades(2000-2020).

3. COMPREHENSIVE ANALYSIS OF THE INTERNATIONAL LANGUAGE TESTING RESEARCH

Although modern language testing research emerged in 1961 and has made great progress in the following decades, statistics of publications shows that there are comparatively fewer research papers before 2008. However, since 2008, the number of publications has increased sharply and there is a continuous growth since then ("Figure 1"). This indicates that language testing has been appealing to more and more scholars and the research has been promoted to a higher level in the past ten years or so.

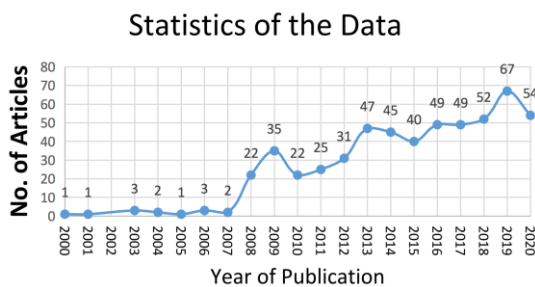


Figure 1 Statistics of the yearly publication.

As for the source journals, language testing and LANGUAGE ASSESSMENT QUARTERLY are far ahead in the number of published articles, 127 and 102 respectively. This is determined by the nature of the two journals. They feature original research on language testing and assessment, covering theoretical issues, empirical studies, and reviews. Thus, these two journals have become the must-read journals and the weather vane of research trends in the field of language testing research. Following the above 2 leading journals are ADVANCES IN SOCIAL SCIENCE, EDUCATION AND HUMANITIES RESEARCH, INTED PROCEEDINGS, FOREIGN LANGUAGE ANNALS and SYSTEM, etc. ("Figure 2") They also publish some articles on language testing research in educational settings.

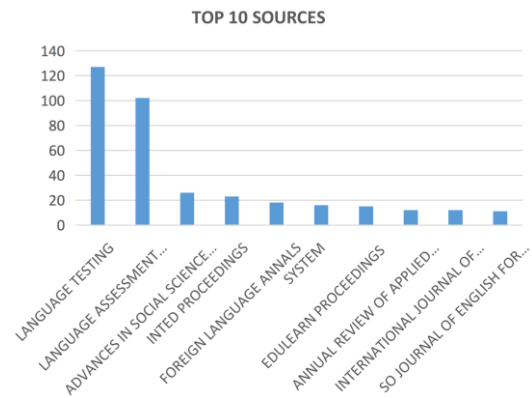


Figure 2 Top 10 article sources.

In terms of countries/region, language testing research is mainly distributed in the United States, the United Kingdom, Australia and China. Among them, American scholars published about 160 articles, which is far more than any other countries. Britain published more than 90 articles. Australia and China published more than 60 articles respectively ("Figure 3"). In recent years, Chinese scholars have paid great attention to publishing in SSCI journals, so that they can enhance their international academic influences. But in any case, America is definitely the leader in both quantity and quality of language testing research.

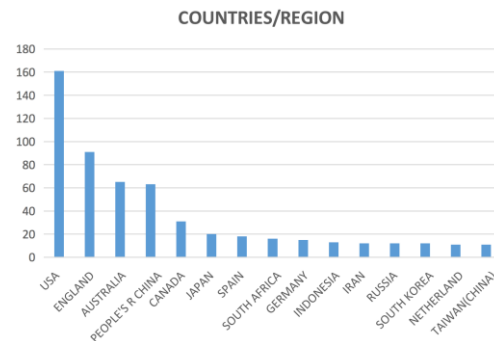


Figure 3 Country distribution of the research.

Finally, as far as the authors are concerned, Tim McNamara from the University of Melbourne in Australia published as many as 14 papers on language testing during 2000-2020. Catherine Elder from America published 8 papers, Constant Leung from UK had 6 publications, John Pill and Carol Chapelle published 5 respectively ("Figure 4"). (Note: The author in this paper refers to the first author or independent author of an article) Their publications are frequently cited documents in the field.

In short, from the international perspective, the year 2008 can be a dividing line in language testing

research. It has developed rapidly since then, more and more researchers, thus more and more research achievements have emerged globally. However, the most outstanding scholars in this field mainly come from the United States, the United Kingdom, and Australia.

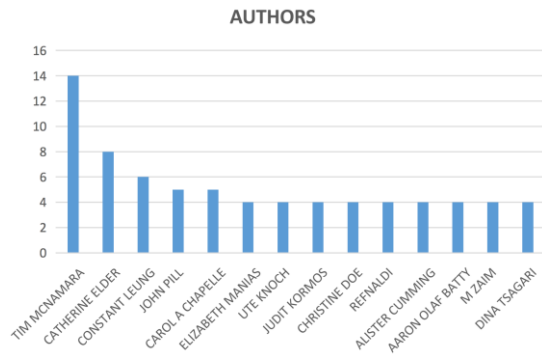


Figure 4 Authors with 4 articles and more.

4. VISUALIZED ANALYSIS OF KEYWORDS

4.1 Keyword Co-occurrence and the Present Research Status

The data is imported into CiteSpace (5.7.R5W) and "Keyword" is selected as the node type. The time slicing is set from Jan, 2000 to Dec, 2020. Pathfinder is selected as the pruning type. The visualized maps are obtained.

"Keywords are the essence and center of the main information of a paper. The analysis of the keywords will accurately reflect the research direction, research hotspots and research trends of the paper" [29]. Apart from the nodes that have no significance in this paper, like "(language) assessment", "language testing" "knowledge", "education" and "language", important large nodes of high betweenness centrality include "English", "proficiency", "performance", "validity", "student", "acquisition", "learner", "comprehension" etc. ("Figure 5"). Betweenness centrality is an indicator to measure the importance of nodes from the network structure. The higher the centrality of a node (generally ≥ 0.1), the stronger the pivot function, and the more information it links between keywords, thus, the more important the position it occupies in the network structure [28]. The nodes of high betweenness centrality suggest that most language testing research in the educational context is involved in English language teaching(efl), mainly about students' language proficiency, performance and acquisition or validity of a certain

testing approach. Specifically, "listening", "reading", "writing", "communication", "discourse", "specific-purpose language" and "bilingualism" are all involved. Besides, research topics concerning language testing itself are also very popular among scholars. In fact, almost every facet concerning testing itself has been studied, including "test taker", "language testers", "test validity", "test score", "high-stake tests", "test item", "test design", "rating scale", "test development", "assessment criterion", "assessment procedure", "assessment literacy", "automated scoring", "construct validation" and "feedback or backwash", etc. Most of the above important keywords have been research topics in the past 20 years or so.

4.2 Keyword Clusters and Main Research Themes

Figure 6 is the visualized map of keyword clusters. In this paper, clusters labels were automatically extracted from indexing terms by the algorithm of LLR. In order to obtain the most typical clusters, small ones with fewer than 10 members were filtered out. Finally, 13 keyword clusters of language testing during the period of 2000-2020 were obtained.

According to Li Jie, et al. (2017), in CiteSpace cluster analysis, if the module value Q is above 0.3, it means that the structure of the division of communities is remarkable. The higher the Q value, the more reliable the network clustering. Besides, Silhouette value S is an index to measure the homogeneity of the network, the closer the S value is to 1, the stronger the homogeneity of the network, thus the more reliable the clustering. If $S > 0.5$, clustering is generally considered reasonable. In "Figure 6", modularity $Q = 0.8165$, weighted Mean Silhouette $S = 0.9327$. Harmonic mean (Q, S) $= 0.8707$. That means the keyword clustering result is very reliable. Some important clusters represent the most noteworthy research themes in language testing [16].

about corpus linguistic methods in language testing research also enrich our understanding of this field. Corpus-based approach has to be given much weight in language testing research.

Online formative assessment (#10) is another newly emerged research theme. For instance, Mohamadi, Z.(2018) compared the effect of online summative and formative assessment on EFL student writing ability [21]. The results implied that using the engaging technology and techniques along with appropriate assessment strategies was a powerful way of making learning efficient. In view of the limited capacity of Dynamic Assessment (DA), Kamrood, A.M., et al. (2019) did a research on diagnosing L2 learners' development through online computerized dynamic assessment, which could help teachers develop fine-tuned individualized learning plans and materials for future learning [13].

Validation (#5) and validity (#13) are also very popular research themes. Fulcher, G., et al. (2011) devised a new scoring instrument called Performance Decision Tree (PDT) for testing speaking after comparing the priorities of the measurement-driven approach and the performance data-driven approach [9]. McNamara, T. & Knoch, U. (2012) examined the uptake of Rasch measurement in language testing [18]. Multi-faceted Rasch measurement makes it possible to address validity issues within performance-based communicative language assessments. Considering the fact that a large number of vocabulary tests were launched with inadequate validation evidence and developers of vocabulary tests did not generally give validation sufficient attention, Schmitt, N., et al.(2020) argued for more rigorous and systematic procedures for test development. They also called for greater assessment literacy among vocabulary test developers [24].

Backwash effect (#7)(S=0.888) is another important research theme in language testing, although it is no longer a very fashionable one. However, positive backwash is recognized as one of the main criteria for evaluating language tests. Additionally, language proficiency (#11) is one of the most enduring research themes in language testing. The map of keyword clusters indicates that language acquisition (#4), integrated tasks (writing) (#6), listening (#9) and reading (#12) are also very significant language testing research areas. Besides, "score fluctuate" and "score use" are appealing to some researchers as well.

4.3 The Diachronic Evolution of Language Testing Research

Visualized cluster map can show the main research themes of language testing, while the timeline atlas of the clusters is designed to display the diachronic evolution of keywords in language testing research ("Figure 7"). It shows that keywords before 2008 are sparser and fewer, indicating that language testing research was not widely distributed among scholars internationally. However, many important keywords emerged before 2008 and still be researched currently, such as the several largest nodes, "proficiency", "English", "performance", "learner" and "student". Then, from 2008, more and more keywords appeared on the timeline in each cluster and more clusters emerged, indicating that language testing is becoming more and more popularly researched in both depth and breadth. Especially, researches concerning student (#1), children (#2) and language proficiency (#11) are the most enduring themes in language testing. And online formative assessment (#10) is the newest one. "Figure 7" indicates that almost all clusters are still very significant research focuses except that backwash effect (#7) and reading (#12) are not as popular as before. All in all, as a research domain beginning to flourish in the past decade, language testing research is bound to develop tremendously in many topics in future.

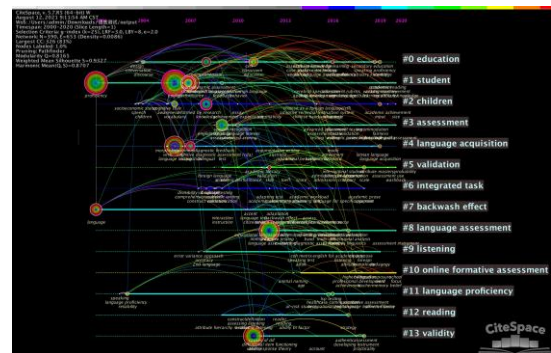


Figure 7 Timeline atlas of keyword clusters.

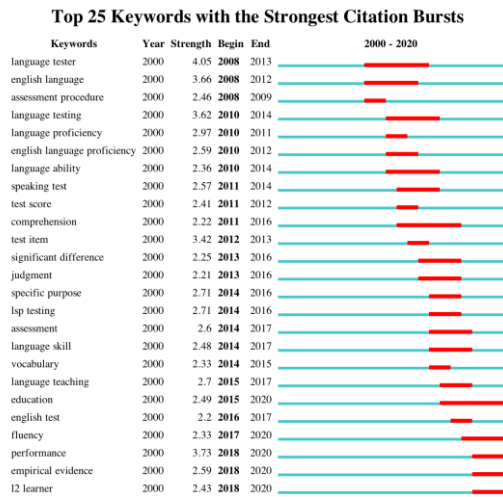


Figure 8 Top 25 keywords.

4.4 Future Forecast of Language Testing Research

Among the plenty of keywords co-occurred in language testing research, there are some enduring and significant ones. They are usually considered to be the lasting or closely followed research focuses. The top 25 keywords with the strongest citation bursts are listed in "Figure 8". Generally, the burst of a keyword can be defined by two attributes: burst duration and burst strength. Unfortunately, there is no keyword continuing bursting throughout our research period (2000-2020). They were changing rapidly and constantly, indicating the rapid changes of research topics. In terms of burst strength, "performance", "fluency", "vocabulary", "judgment", "construct validation", "speech", "proficiency", "validity", "accuracy", "learner", "teacher", "reading" and "cefr" are stronger ones (strength > 2.00), except "language testing" and "(language) assessment". Then, "fluency", "achievement", "performance", "teacher" and "cefr" are the newly emerged keywords of strong citation bursts that last till now. Research involving these keywords are probably hotspots in language testing at present.

Compared with the hotspot analysis, the future forecast of research frontiers is more valuable and directive for language testing researchers. The Sigma values of all the following keywords are above 1.0: "speech", "proficiency", "performance", "learner", "construct validation", "accuracy", "validity". The higher the Sigma value of a certain node is, the more possible it becomes a research frontier. It is predictable that topics involving the above keywords are likely to be still research

hotspots in the near future. They may be researched from different perspectives or with new methods and theories.

5. MOST HIGHLY CITED AUTHORS AND REFERENCES

With the flourishing of language testing research, more and more influential authors and literature emerged, particularly in the past two decades. Table 1 shows that Bachman, L.F., Alderson, J.C., McManara, T. are the top three most frequently cited authors in the past two decades. Bachman L.F. is the most highly cited author. His publications were cited 198 times in the past 20years. And he is also the author with the strongest citation burst. Alderson J.C. is another very productive author with very high citations (132 in our research period) in the field of language testing. Alderson, J.C., et al (2013) questioned the use of expert judgments and their usefulness in distinguishing between construct components [1]. He (2015) proposed a set of five tentative principles of diagnostic language assessment that might be important references for further research [2].

Another highly cited author Mcnamara, T. (2000) gave a succinct theoretical introduction to the basic concepts in language testing. His comments on the social aspects of language testing are so important to us in designing language test items [19]. Then, he challenged the authority of Common European Framework of Reference for Languages (CEFR) in the determination of test constructs [20]. On the contrary, he presented a comprehensive framework for researching classroom-based assessment (CBA) processes, based on a detailed empirical study.

Besides them, Shohamy, E., Fulcher, G., Davies, A., Kane, M.T. and Council of Europe, etc., were also cited very frequently in our research period. All of these authors are leading figures in the field of language testing. However, in recent 5 years, Ginther, A, Chapelle, C.A., Hyland, K., Xi, X., Pill, J. and Kormos, J., etc., had very strong citation bursts. Esp., Ginther, A, Pill, J. and Kormos, J. maintain very strong citation bursts till the present time. Their research probably relates to the current hotspots in the field of language testing.

As far as the important references are concerned, Bachman L.F. (2010) and McNamara T. (2006) are the first two in terms of citation burst ("Figure 9"). As two of the greatest scholars in the field of language testing, their publications are classic

documents for us to refer to. Additionally, Kane, M.T. (2013), Tailer, L. (2009), Shohamy, E. (2011) are also documents which have very strong citation bursts. Particularly, the citation bursts of the last 5 documents, namely American Educational Research Association, American Psychological Association & National Council for Assessment in Education (2014), Kane, M.T. (2013), Shohamy, E. (2011), Pill, J. (2013) and Fulcher, G. (2012), continue to be very strong till now. The ideas or methodology in them are likely to be the newest or the most advanced, denoting the present hotspots in language testing research.

Table 1. Highly cited authors (2000-2020)

Author	Freq	Author	Freq
BACHMAN LF	198	Brown A	42
ALDERSON JC	132	CUMMING A	42
McNamara T	116	Douglas D	42
Shohamy E	82	Weigle SC	41
FULCHER G	79	Weir CJ	41
Davies A	76	Canale M	36
Kane MT	72	Buck G	33
Council of Europe	69	Kane M	33
Messick S	69	Chalhoub-Deville M	32
Chapelle CA	60	DAVIDSON F	32
Elder C	60	KNOCH U	30
TAYLOR L	54		

"teacher" and "cefr" are likely to be hotspots in language testing currently and in the future. Bachman, L.F., Alderson, J.C., McManara, T. are the most frequently cited authors in the past two decades, while Bachman L.F. (2010) and McNamara T. (2006) keeps the strongest citation burst. However, the authors and references with very strong citation bursts in recent 5 years probably represent the new research trends in language testing. With the emergence of new research methods and theories, language testing research will develop more prosperously in future.

AUTHORS' CONTRIBUTIONS

Yansong Feng is responsible for the data collecting and paper writing; Jianghong Han analyzes the data and contributes to the revising and editing of the paper.

ACKNOWLEDGMENTS

This research was financially supported by Jiangsu Provincial Social Science Foundation Project. (Project Name: Research on Language Testing at Home and Abroad; Grant Number: 12YYD019).

REFERENCES

[1] Alderson, J.C.; Kremmel, B., 2013, Re-examining the content validation of a grammar test: The (im)possibility of distinguishing vocabulary and structural knowledge, LANGUAGE TESTING, V30, P535, DOI: 10.1177/0265532213489568

[2] Alderson, J.C., 2015, Towards a Theory of Diagnosis in Second and Foreign Language Assessment: Insights from Professional Practice Across Diverse Fields, APPLIED LINGUISTICS, V36, P236, DOI: 10.1093/applin/amt046

[3] Barbosa, Poliana, G., 2019, The role of working and short-term memory in predicting receptive vocabulary in monolingual and sequential bilingual children, INTERNATIONAL JOURNAL OF BILINGUAL EDUCATION AND BILINGUALISM, V22, P801, DOI 10.1080/13670050.2017.1314445

[4] Bernstein, Jared., 2010, Validating automated speaking tests, LANGUAGE TESTING, V27, P355, DOI 10.1177/0265532210364404

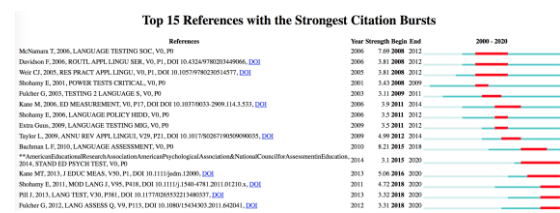


Figure 9 Top 15 references.

6. CONCLUSION

Using CiteSpace as the research tool, this paper gives a comprehensive analysis of the development course of the international language testing research during the past 2 decades from five aspects: keyword clusters, timeline atlas, keyword co-occurrence, highly cited authors and references. The study shows that the development of language testing research in the past two decades can be divided into two stages with the year 2008 being the dividing line. The leading figures in this field are mainly from USA, UK and Australia. Besides the research themes concerning students' or children's language learning in educational settings, research about test validity or online formative assessment is also very popular at present. Research involving "fluency", "achievement", "performance",

- [5] Brooks, F.B. & Darhower, M.A., 2014, It Takes a Department! A Study of the Culture of Proficiency in Three Successful Foreign Language Teacher Education Programs, *FOREIGN LANGUAGE ANNALS*, V47, P592, DOI 10.1111/flan.12118
- [6] Castillejo, Susana Perez, 2019, The role of foreign language anxiety on L2 utterance fluency during a final exam, *LANGUAGE TESTING*, V36, P327, DOI 10.1177/0265532218777783
- [7] Davies, A., 2014, Fifty years of language assessment[A]. In Kunnan A J (Ed.) . *The Companion to Language Assessment Vol. I* [C]. Chichester: Wiley Blackwell, pp.3-16.
- [8] Egbert, Jesse., 2017, Corpus linguistics and language testing: Navigating uncharted waters, *LANGUAGE TESTING*, V34, P555, DOI 10.1177/0265532217713045
- [9] Fulcher, G., Davidson, F. & Kemp, J., 2011, Effective rating scale development for speaking tests: Performance decision trees, *LANGUAGE TESTING*, V28, P5, DOI 10.1177/0265532209359514
- [10] Hill, K. & McNamara, T., 2012, Developing a comprehensive, empirically based research framework for classroom-based assessment, *LANGUAGE TESTING*, V29, P 395, DOI 10.1177/0265532211428317
- [11] Hsieh, Mingchuan, 2013, An application of Multifaceted Rasch measurement in the Yes/No Angoff standard setting procedure, *LANGUAGE TESTING*, V30, P491, DOI 10.1177/0265532213476259
- [12] JIN Yanhua, 2020, Development trend, hot spots and enlightenment of international language testing research:Based on bibliometric analysis (2008-2019), *Journal of Zhejiang Sci-Tech University*, 44(3):227-235, DOI:10.3969/ji.sn.1673-3851(s).2020.03.001
- [13] Kamrood, A.M., Davoudi, M., Ghaniabadi, S., Amirian, S.M.R. 2019, Diagnosing L2 learners' development through online computerized dynamic assessment. *COMPUTER ASSISTED LANGUAGE LEARNING*, V7, P30, DOI 10.1080/09588221.2019.1645181
- [14] Kim, Youb, 2008, The effects of integrated language-based instruction in elementary ESL learning, *MODERN LANGUAGE JOURNAL*, V92, P431, DOI 10.1111/j.1540-4781.2008.00756.x
- [15] LaFlair, G.T. & Staples, S., 2017, Using corpus linguistics to examine the extrapolation inference in the validity argument for a high-stakes speaking assessment, *LANGUAGE TESTING*, V34, P451, DOI 10.1177/0265532217713951
- [16] Li Jie, Chen Chaomei. *Cite Space: Scientific and Technological Text Mining and Visualization (Second Edition)* [M]. Beijing: Capital University of Economics and Business Press, 2017: 77-171.
- [17] Liu Zeyuan, Chen Yue, Hou Haiyan, etc. *Method and Application of Scientific Knowledge Map* [M]. Beijing: People 's Publishing House, 2008: 4.
- [18] McNamara, T. & Knoch, U. 2012, The Rasch wars: The emergence of Rasch measurement in language testing. *LANGUAGE TESTING*, V29, P555, DOI 10.1177/0265532211430367
- [19] Mcnamara, T., 2000, *Language Testing*, Oxford: Oxford University Press.
- [20] Mcnamara, T., 2011, Managing learning: Authority and language assessment, *LANGUAGE TEACHING*, V44, P500, DOI:10.1017/S0261444811000073
- [21] Mohamadi, Z., 2018, Comparative effect of online summative and formative assessment on EFL student writing ability, *STUDIES IN EDUCATIONAL EVALUATION*, V59, P29, DOI 10.1016/j.stueduc.2018.02.003
- [22] Park, K., 2014, Corpora and Language Assessment: The State of the Art, *LANGUAGE ASSESSMENT QUARTERLY*, V11, P27, DOI 10.1080/15434303.2013.872647
- [23] Sandilos, Lia E, 2015, Analysis of Bilingual Children's Performance on the English and Spanish Versions of the Woodcock-Munoz Language Survey-R (WMLS-R), *LANGUAGE ASSESSMENT QUARTERLY*, V12, P386, DOI 10.1080/15434303.2015.1100198

- [24] Schmitt, N., Nation, P. & Kremmel, B., 2020, Moving the field of vocabulary assessment forward: The need for more rigorous test development and validation, *LANG TEACHING*, V53, P109, DOI 10.1017/S0261444819000326
- [25] Weigle, Sara Cushing., 2010, Validation of automated scores of TOEFL iBT tasks against non-test indicators of writing ability, *LANGUAGE TESTING*, V27, P335, DOI 10.1177/0265532210364406
- [26] Wu, Siew Mei., 2016, Managing rater effects through the use of FACETS analysis: the case of a university placement test, *HIGHER EDUCATION RESEARCH & DEVELOPMENT*, V35, P380, DOI 10.1080/07294360.2015.1087381
- [27] Xi, Xiaoming, 2017, What does corpus linguistics have to offer to language assessment? *LANGUAGE TESTING* V34, P565 DOI 10.1177/0265532217720956
- [28] Yansong Feng. A Diachronic Research of the Development of Foreign Linguistics in China--A Knowledge Map Analysis Based on the National Social Science Fund Project Database (1996-2019). Proceedings of the 7th International Conference on Education, Language, Art and Inter-cultural Communication (ICELAIC 2020). V507, P195.
- [29] Yuan Zhoumin, Liu Huanhuan. Dynamic Visual Analysis of International Interlanguage Pragmatics Research [J]. *Foreign Language Teaching Research*, 2017 (5): 456-463.
- [30] Zhou Shanshan. A State-of-the-art Review on Focused Topics of International Language Testing (2011-2015) *Foreign Language Testing and Teaching*. 2018,(02):1-14.