

A Developmental Study on Modelling Proficiency Test of an Indigenous Language in Indonesia

Usep Kuswari

Univeritas Pendidikan Indonesia
usep.kuswari@upi.edu

Abstract

The paper considers that the necessity of establishing proficiency test of an indigenous language, namely Uji Kemahiran Berbahasa Sunda (UKBS) is indispensable. Guided by the fact that some languages have established their official proficiency test, UKBS has to devise a way to compose its competence map, general test material, and material selection. The present study mainly combines the respective model exemplified by Borg and Gall (1983) and Kirkpatrick (1998). The model covers initial stage, designing stage, trial (revisionary) stage, and practical implementation stage. The study only focuses on the initial stage and designing stage. UKBS has sections (girangan) and materials. UKBS encompasses five pinpoint topics, namely girangan I: ngaregepkeun (listening), girangan II: ngarespon kaidah tata basa Sunda (responding Sundanese grammar), girangan III: maca (reading), girangan IV: nulis (writing), and girangan V: nyarita (speaking). It is a paradigm that UKBS is a proficiency test, not an achievement test. Therefore, the test refers to actual language use situation. The topics have a way of alluding to culture, social, education, profession, and scientific realm. Eventually, the validity test proves that all test items for each section are reliable. Nevertheless, the findings should be integrated to trial and practical implementation stage for the betterment of UKBS model.

Keywords: UKBS, Competence map, Proficiency test, Sundanese language, Language use

1. Introduction

In learning process, evaluation has an important role in identifying students' competence. Evaluation displays considerable empirical evidence of the process. By having the evidence, teacher and related parties can use it for the betterment of upcoming activity. In this sense, an evaluation needs to be accompanied by well-planned, well-arranged, and well-implemented instruments. One of which is test. Language teaching uses test as an evaluation tool to identify language proficiency. Some languages have established their official proficiency test, namely TOEFL, DALF, TOAFL, IMALAH, and B2GER. Despite of the fact, Sundanese, as an indigenous language with great number of speakers in Indonesia, has not established a reliable tool yet. Therefore, the necessity of establishing Sundanese proficiency test, namely Uji Kemahiran Berbahasa Sunda (UKBS) is indispensable.

In relation to test, Ellis in [1] states that all tests are formal assessment but not all formal assessments are test. In general, evaluation can be defined as a collection of information about the qualitative or quantitative improvement of students' learning outcomes. Nokelainen in [2], argue that assessment has objectives to (1) determine the level of students' knowledge and skills,

Article history:

Received (February 4, 2020), Review Result (March 6, 2020), Accepted (April 12, 2020)

(2) capture the students' improvement to be implemented in hereinafter teaching program, and (3) provide data to measure the final level of students' outcome. Moreover, O'Mally and Pierce (1996, p.20) suggest that the quality of language learning evaluation is determined by clear objectives so as to make the assessor receives effective responses and sincerity from students. The selection and the use of proper evaluation approach are useful to discover language competence and performance. In this sense, (a) management of assessment tools (test and non-test classroom information, class record, and test record), (b) presentation of results (in numbers) and truths (in percentages), and (c) utilization of test results are indispensable resources in the development of assessment instrument.

In the test implementation, input needs to take into consideration. Input, in this context, is related to two aspects, namely format and nature of language. The format includes channels and presentation forms, presentation tools, presentation language, problems identification, and speed level. The input can be presented orally or visually in receptive mode, whereas answer can be either spoken or written in productive mode. It is a fact that Sundanese proficiency test is developed by disaggregated approach. This approach focuses on language theory and behaviouristic theory. It tends to assess linguistic competence in theoretical viewpoint. However, most of curricula say that proficiency test is intended to measure language competence and language performance in natural context of language use.

The inability to manage Sundanese proficiency test also makes ineffective learning process. The situation produces low performance of Sundanese proficiency level. Therefore, this research is obliged to fill the gap of Sundanese proficiency test model. After doing so, the research hopes that the model of proficiency test can be followed by the development of Sundanese learning model. The next session will elaborate how Sundanese proficiency test is constructed.

The rest of this paper is organized as follow: Section 2 describes the proposed literature review. Section 3 describes the proposed method. Section 4 describes the findings. Finally, Section 5 concludes this work and following by pedagogical implication.

2. Literature review

2.1. Evaluation defined

Making definition of a literary term is always a labyrinthine business. Evaluation is a systematic process of determining the extent to which educational objectives are achieved [3]. The definition covers two concepts; (a) systematic process is related to series activities that conform to rules and (b) evaluation is always connected to pre-determined teaching objectives. These objectives are used as guidelines to develop evaluation instruments.

Menges in [4], produced a series of arguments regarding evaluation. Menges in [4] stated that evaluation is a systematic process of determining the level of achievement and/or objectives of learning materials received by students. Subsequently, Menges in [4] defined evaluation as a systematic process for collecting and interpreting information about learning achievement, then to determine students' score. Furthermore, Stufflebeam and Shrinkfield in [5], strengthened that evaluation is a systematic assessment of the worth or merit of some objects. Stufflebeam in [5] further explained that evaluation is "an ongoing check on the implementation of a plan" (p. 174). It confirms that evaluation is a systemic process from initial to final activities. In addition, Lynch in [6] supported the previous statement by stating that evaluation is a systematic attempt to gather information in order to make judgements or

decisions. However, Troudi in [7] considered that the concept of evaluation is broader than the concept of judgement.

Evaluation often follows measurement; taking the form of judgement about the quality of a performance [8]. This statement shows that the ultimate focus on evaluation is performance quality. However, 'evaluation' (as a term) makes confusion over its use in learning process [9]. In practice setting, 'evaluation' is synonymous with 'measurement' or 'testing'. It refers to 'achievement test', 'achievement measurement', or 'achievement evaluation'. In different context, 'evaluation' is also defined as a research method, which is independent of measurement tools.

A test can be defined as an instrument of judgement. On the other hand, a test can also be defined as a measurement technique. As stated by Ellis in [1], a test will be defined as a systematic procedure for measuring a sample of an individual's behaviour. Based on the definition, a test has two important aspects, (1) systematic procedure and (2) measuring specific sample of behaviour. 'Systematic procedure' means that a test must be organized, processed (managed), and implemented based on certain rules. Moreover, 'systematic' reflects two meanings (a) systematic in its content; a test items must be organized and selected based on the area of behaviour that will be measured, so that the test meets its validity, (b) systematic in its implementation (administration); a test must be implemented (managed) based on certain rules, and (c) systematic in its data processing; it means that the data generated from a test are processed and interpreted based on certain rules. The term 'measuring of an individual's behaviour' shows that a test only measures a specific sample of behaviour. A test cannot measure the entire population of behaviour, but is limited to the specific test item.

A test, which has question that must be answered individually, is defined as achievement test. Hickey *et al.* in [10], in relation to achievement test, stated that the type of ability test that describes what a person has learned to do is called an achievement test. This approach is based on structuralism theory of language, which sees that language is a collection of discrete elements, and formed by a certain structure. In addition, this approach is also supported by psychometric viewpoint that allows each element or unit of language to be measured separately.

The respective approach has three advantages. The advantages are (a) it is easy to be quantified because the evaluation is done on atomistic and discrete parts, (b) the test items (of this approach) can be much wider or more encompassing various levels and language skills, and (c) it meets cost-efficiency and intelligibility in administration. In addition, this approach also produces quantifiable data of broad topics of language (phonology, morphology, syntax, and vocabulary), which are measured. The 'discrete' approach that focuses on linguistic topics is efficient and has general reliability in test achievement. However, the two approaches (and formats) used have disadvantages in measuring process.

2.2. Proficiency test

Most of scholars would probably agree that measuring linguistic comprehension is necessary but not sufficient to display language proficiency. van Gelderen in [11] criticized the test that mainly focuses on linguistic aspects. Stern in [12] stated that grammatical comprehension is not related to communicative skill. Littlewood in [13] argued that the purpose of applying linguistic test is to apply linguistic analysis skill. For example, it is a comparison between the introduction of context-specific meaning of utterances and system-giving meaning of utterances. To talk about this analysis, the language transfer should be more reliable for skill's measurement and development. The language transfer is important to establish communicative

achievement and production (and comprehension) of integrated lectures. Those aspects are more important than linguistic topics.

O'Donoghue in [14], explained that measuring language proficiency should be directed to the use of language. The statement reflects a general concept that language proficiency is an amalgam of knowledge and ability to perform language structures. A test, therefore, should consider context in language use rather than knowledge of language elements. It is also important to highlight the shortcomings of psychometric-structuralism approach. The shortcomings are related to the facts that (a) separate testing (of discrete elements) does not pay much attention to interaction between all language elements in wider and more complex communication context, (b) the test makes an ineffective assessment as the main part of language disappears, (c) it would be unwise to predicate that grammatical competence reflects communication skill, and (d) the approach is artificial, sterile, and irrelevant as language performance is assessed in unreal context of situation. Melitz in [15], emphasized that integrative test is better than single or combination test in presenting all language skills. Cloze and dictation become highly integrative because they fill all language skills. The focus is on high correlation between cloze and another dimension. Each individual has different skill. However, it does not mean that there will be no individual achievement in all skills.

O'Donoghue in [14], provided a fact that cloze or dictation does not offer an opportunity, for students and examiners, to produce language rules spontaneously. The test procedure does not offer possibility of written or spoken production as it is usually held for a very important communicative situation. Although integrative measures appear to relate similar measurement of general language proficiency, there is a general statement on the cloze associated with a written production test. The associated test is trustworthy and states that the probability of this area of expertise cannot be adequately predicted by an overall skills test.

2.3. Communicative language skill

Lynch in [16] argued that, before having a test, providing task and particular training for students is important. Analysing students' answers, in a test, cannot directly provide relevant information about students' difficulties in an original task. Therefore, an indirect test is reliable and provides legitimate measurement. By doing indirect test, the evidence of achievement levels can be shown but cannot diagnose specific areas of difficulty in relation to the original task. The approach is based on communicative competence. The competence should reflect in daily life communication. Therefore, language skills test, in this approach, measures capacity McCormick in [17], communicative language proficiency, evaluating samples of performance, in certain specific contexts of use, created under particular test constraints [8].

Jcgm in [8], explained that communicative language skills cover language competence and capacity to use the competence in language use. The essence of communicative language skills test is a single process to evade juxtaposing different language competence and capacity McCormick in [17] or communicative language capacity [8]. All juxtaposing actions cannot represent actual capacity in using language. On the contrary, capacity is related to evaluation of understanding in language use through its embodiment. Achievement can only be directly observed; therefore, evaluation is important to do. O'Donoghue in [14] stated that there should be a distinction between communicative competence and communicative achievement as distinctive feature is a fact that achievement is the embodiment. This confirms what Celce-Murcia in [18] stated about the three capacities and their interaction.

Communicative language concerns language format, how to use language in proper context, and students' understanding level in performing language [14], [18]. In different way,

Machado-da-Silva in [19] stated that competence of communication with pleasure influences sociolinguistic determination. Hudson in [20], strengthened that the primary need of students is not analytical and theoretical competence of target language, but the ability to understand and be understood in a certain context of language use. Fetterman in [19], in another attempt, argued that all tests can be used to see the degree of communicative or non-communicative language use and its basic use. Furthermore, tests also need to consider a subject as meaning-dependent and describe the preceding test to the involvement of a context in making (or not making) response. To take part in a communicative event, students are directed to produce and understand a lecture in a certain context-situation, and then students make a link to the acquired attainment. The purpose of this test is to assess students' capacity in taking part in communicative event.

In defining test questions description, the writing style of related question in target language should be similar to language activity in similar situation. Moreover, examiners are allowed to make measurement planning of their performance assessment that cannot be separated from their own task. The strength of test questions form is useful by considering test size (text length, receptive or productive involvement), the grammar complexity and the scope of required cohesion tools, functional coverage (the level of illusionary variation), and instruction coverage (the relation between knowledge and required activity).

In different sense, speaking test (in proficiency test) is not only about oral test but also performance test (non-verbal test). This means that evaluation is not only derived from the act of speaking but also the process of conversation production. Therefore, speaking test techniques are aided by observation techniques; examiners observe (not only listen) how students speak. This applies to direct oral performance testing. To reveal the ability to speak Sundanese, images can be a decent stimulus of speaking. The use of that stimulus is relevant to the speaking assessment of elementary school students. However, the image stimulus can also be applicable for higher-level students by conforming image complexity and students' capacity. According to Tomblin in [21], decent images should make students eager to speak or perform their idea in proper language. The tasks assigned to the students can be in the form of giving questions and telling stories.

Based on scientific point of view, this research is an applied research, which focuses on applying language appraisal (speaking and writing) in language teaching and learning. This research uses various theories on Sundanese speaking and writing test and its relation to draw a firm conclusion. As an instrument for measuring language skills, language judgments must not only meet certain criteria, but also should be based on a particular theoretical framework. Theoretically, language can be seen as a system and an action. Structuralism sees language as a system of arbitrary symbols used for human communication. Therefore, scholars believe that language is structured by some small parts; therefore, each part can be acquired separately. Sundanese speaking and writing test employs an integrative approach. Speaking test can be formulated in some types. The examples are (a) verbal essay (students are instructed to speak for three minutes in a general topic (or more), (b) oral presentation (students is expected to speak briefly on a prepared topic), (c) free interview (students make conversation in unstructured model and unsettled procedures, (d) controlled interview, and (e) role play (students play one role in an interaction setting that may happen in real situation).

3. Proposed method

The research method used in this research is a development model by using research and development research type. This aims to produce a product in the form of Sundanese

proficiency test model on students of UPI. Educational research and development is a type of research that is widely used to solve practical problems in education. As stated by Borg and Bennett in [22], research and education development are a process used to develop and validate educational products.

In this study, inductive and deductive procedures are used to reveal the meaning of obtained data. This confirms Troudi in [7] who states that two procedures open to researchers are inductivism and deductivism. Therefore, the procedures of developing Sundanese learning model are as follows.

- a. Research and information collection (including literature review, classroom observation, and report preparation).
- b. Planning (including defining skills, declaring objectives that determine the order of the series, and the opportunity of small-scale trials).
- c. Developing a preliminary pattern of the product (including learner material, manuals, and evaluation tools).

Based on the ten development procedures proposed by Bennett [22], The procedures can be combined with the McFarlane in [23] model and, finally, can produce the UKBS model. The procedures proposed by Bennett in [22] are relatively identical to Kirkpatrick model [23]. Therefore, this study simplifies these steps into the development procedure of UKBS model. The procedures are: (1) the initial stage; (2) the design stage; (3) trial and revision stage; and (4) implementation stage. However, in this study, there are only two stages taken into account, namely (1) the initial stage and (2) the design stage.

4. Findings

The results of this study include (1) the Sundanese competence map: *ngaregepkeun* (listening), responding to the Sundanese grammar, *maca* (reading), *nulis* (writing) and *nyarita* (speaking) in the UKBS; (2) materials of UKBS; and (3) UKBS guidelines. Competence map of *ngaregepkeun* (listening) aims to mark the proficiency of obtaining information orally, both in the form of dialogue and monologue. The proficiency of *ngaregepkeun* (listening) is a skill that aims to measure listening skill of participants. Participants are exposed to dialogue or monologue. In the proficiency map of *ngaregepkeun*, there are four dialogue texts and four monologues. In accordance to the label, the materials are delivered orally and accepted by students through the means of hearing. The immediate problem is what media to use, should the teacher use the recording media or directly deliver it orally during the test.

The ability of listening can be interpreted as the ability to capture and understand other languages. Therefore, the appropriate linguistic material is certainly a discourse, since a discourse must contain real information. For tests of listening competence, the selection of test subjects is more emphasized on the state of discourse, both in terms of difficulty level, content, and types of discourse. The level of discourse's difficulty is mainly observed from the vocabulary and the structure used. If the vocabulary used is difficult, ambiguous, and abstract, rarely used, and complex-structure, the discourse will be in high-level discourse. A reliable discourse to be used in listening skill test is rather a mid-level of discourse.

The content and scope of discourse also affect the level of discourse difficulty. If the content of the discourse is not relevant to participants' interest and need, or not relevant to what participants learn, it will increase the level of respective discourse difficulty. The discourse should have neutral information, so it is possible to have common sense of the problem. Instead, it is not relevant to use a discourse that contains a certain group's view or belief because it will lead to dissent view.

The competence map of *ngaregeokeun* skill can be seen in the following [Table 1].

Table 1. The competence map of *ngaregeokeun* skill

N O.	TYPES OF DISCOURSE	THEME	TITLE
1	Monologue	Culture	<i>Adat Istiadat dina Ngalamar</i>
		Sport	<i>Piala Dunya Mugi Janten Conto</i>
		Social	<i>Hukuman</i>
		Religion	<i>Dyukur Ni'mat</i>
2	Dialogue	Culture	<i>Latihan Kabaret</i>
		Sport	<i>Badminton</i>
		Social	<i>Aki-Aki Hade Budi</i>
		Religion	<i>Indung jeung Anak</i>

The competence map of responding Sundanese grammar is the participant's skill in understanding the error of Sundanese structure. In each item, there are two parts in bold and underlined form. Participants are asked to respond to the two sections. The wrong part is replaced by an answer which is provided in below part, then, the correct part is ignored. The most important competence are grammatical structures and Sundanese vocabulary. The grammar includes morphological and syntactic issues, either separately or integratively. The material of grammatical competence should be the representative of the explained material or reflect the purpose of the test.

Table 2. The competence map of responding Sundanese grammar

NO	SUNDANESE GRAMMAR	LANGUAGE ERROR
1	Morphology	The use of word type
		The use of <i>kecap rundayan</i>
		The use of <i>kecap rajekan</i>
		The use of <i>kecap kantetan</i>
2	Syntax	The use of <i>frasa</i>
		The use of <i>klausa</i>
		The use of <i>wangun kalimat</i>
		The use of <i>jenis kalimat</i>
		The use of <i>kalimah efektif</i>
3	Vocabulary (<i>Kandaga Kecap</i>)	The use of <i>kecap umum</i>
		The use of <i>kecap husus</i>
		The use of <i>istilah</i>
		The use of <i>idiom</i>

Based on the [Table 2] above. The competence of *maca* (reading) is related to reading comprehension. The competence map of *maca* (reading) sees this reading ability from the way students process a discourse in three main abilities: (a) the ability to seek and find information (retrieving information); (2) the ability to develop meaning derived from the information, find and make inferences by using one or more information, and (3) the ability to reflect and evaluate the content of the discourse in relation to daily experience, prior knowledge, and the development of ideas from the obtained information.

The context of reading is related to the purpose of the discourse preparation. The contexts include cover: (1) reading for personal interest, (2) reading for public purposes, (3) reading for work purposes, (4) reading for academic purposes

The competence map of *maca* skill can be seen in the following [Table 3].

Tabel 3. The competence map of maca skill

NO.	THEME	TITLE
1	Culture	<i>Ngabuburit</i>
	Sport	<i>Penca Silat</i>
	Social	<i>Midangdam Jurnalisme Ngatik Ngadidik</i>
	Religion	<i>Kakuatan jeung Kamampuh Diri</i>
	Technology	<i>Nyieun Batre tina Apel</i>

The competence of nulis (writing) is a test phase to examine Sundanese speakers' skills in delivering written ideas by using guided techniques. In each test item, there are sentences of plot and pictures, diagrams, or tables that support the information of the discourse. The test participants are asked to describe and develop the sentence in writing product for, at least, 200 words.

The competence of nyarita (speaking) is a test phase to examine Sundanese speakers' skill in conveying ideas orally by using guided techniques. In each test item, there are sentences of scribble and pictures, schematics, or charts that support the information of the discourse. The test participants are asked to describe and develop the in oral product within 10 minutes. The Sundanese Proficiency Test (UKBS) is a proficiency test, not an achievement test. Language proficiency test refers to Sundanese language usage criteria (actual language usage situation). The criteria refer to general life skills, i.e. cultural, social, technological, educational, and special skills areas, namely the domain of professionalism and the scientific realm.

Similar to TOEFL, UKBS also consists of several different sessions and materials. UKBS is divided into four stages or sessions (*girangan*): *Girangan I*: ngaregepkeun, *Girangan II*: Ngarespons kaidah gramatikan basa Sunda, *Girangan III*: maca, *Girangan IV*: nulis, and *Girangan V*: nyarita.

For more details, the standard of Sundanese competence material can be seen in the following [Table 4].

Table 4. The material of UKBS

SesiSession (<i>Girangan</i>)	Sundanese Competence	Material	Time
I	<i>Ngaregepkeun</i> (Listening)	Listening to 5 discourse of monologues and dialogues	30 minutes
II	Responding Sundanese Grammar	Findings language errors of written text	20 minutes
III	<i>Maca</i> (reading)	Reading 5 written texts	45 minutes
IV	<i>Nulis</i> (writing)	Asking for presenting picture/diagram/table in written form within 200 words	30 minutes
V	<i>Nyarita</i> (speaking)	Asking for presenting picture/diagram/table in written form within 200 words	15 minutes

Based on the results of competence mapping, the selection of UKBS material can be seen in the following [Table 5].

Table 5. The selection of UKBS material

SesiSession (Girangan)	Item	Time	Material
<i>Girangan I: Garegepekeun</i>	40 items	30 minutes	Spoken discourse in the form of 4 dialogues and 4 monologues in which 4 topics
<i>Girangan II: Responding Sundanese grammar</i>	40 items	30 minutes	Written test items in which participants select alternative choices from inaccurate part
<i>Girangan III: Maca (reading)</i>	40 items	45 minutes	Written discourse in the form of 5 texts within 5 topics (each text has 8 test items)
<i>Girangan IV: Nulis (writing)</i>	1 test item	30 minutes	Written test items that ask participants to present picture/graph/table in the form of written discourse within 200 words
<i>Girangan V: Nyarita (speaking)</i>	1 test item	15 minutes	Written test items that ask participants to present picture/graph/table in the form of spoken discourse within 5 minutes talk

5. Conclusion and pedagogical implication

Evidently, the necessity of establishing Sundanese proficiency test, namely *Uji Kemahiran Berbahasa Sunda* (UKBS) is indispensable. Guided by the fact that some languages have established their official proficiency test, this study has compose competence map, general test material, and material selection for UKBS. The competency map of the UKBS includes five points, namely (1) competence of listening, (2) competence of responding Sundanese grammar, (3) competence of reading, (4) competence of writing, and (5) competence of speaking.

Moreover, UKBS is a proficiency test, not an achievement test. Language proficiency test refers to Sundanese language usage criteria (actual language usage situation) faced by participants. These include general life skills, i.e. cultural, social, technological, educational, and special skills areas, namely the domain of professionalism and the scientific realm. As similar to TOEFL, UKBS also consists of several different sessions and materials. UKBS material is divided into four stages or sessions (girangan): Girangan I: ngaregepekeun, Girangan II: Ngarespons kaidah tata basa Sunda, Girangan III: maca, Girangan IV: writing, and Girangan V: nyarita.

Statistically, the feasibility of UKBS is Listening test item of UKBS meet fair-validity. That is $r_{count} (0.8) > r_{table} (0.207)$ in the free degree of $p < 0,5$. Therefore, it can be concluded that the UKBS listening test item is reliable. In relation to responding Sundanese grammar, the test item of UKBS meet fair-validity. That is $r_{count} (0.6) > r_{table} (0.294)$ in the free degree of $p < 0,5$. Therefore, it can be concluded that the UKBS responding Sundanese grammar test item is reliable. The reading test item of UKBS meet fair-validity. That is $r_{count} (0.8) > r_{table} (0.207)$ in the free degree of $p < 0,5$. Therefore, it can be concluded that the UKBS reading test item is reliable. The writing test item of UKBS meet fair-validity. That is $r_{count} (0.8) > r_{table} (0.294)$ in the free degree of $p < 0,5$. Therefore, it can be concluded that the UKBS writing test item is reliable. The UKBS nyarita (speaking) test item meet fair validity. It can be seen that factor load value is $(\lambda) > 0.3$, P-value is 0.965, and Chie-Square is 0.029 in df 1.03. These results illustrate that there is a firm relation between the instruments of speaking test and speaking proficiency.

References

- [1] R. Ellis, "Language teaching research and language pedagogy," John Wiley & Sons, (2012)
- [2] P. Nokelainen, "An empirical assessment of pedagogical usability criteria for digital learning material with elementary school students," *Educ. Technol. Soc.*, vol. 9, no.2, pp.178-197, (2006)
- [3] R. a Berk, "Survey of 12 Strategies to Measure Teaching Effectiveness," *Int. J. Teach. Learn. High. Educ.*, vol. 17, no.1, pp.48-62, (2005)
- [4] R. J. Menges and B. G. Davis, "Tools for Teaching,," *J. Higher Educ.*, vol. 66, no.2, (1995) DOI: 10.2307/2943914
- [5] D. L. Stufflebeam and A. J. C. N.-E. 371. 2. S. T. U. N. O. T. O. N. L. Shinkfield, "Systematic evaluation : a self-instructional guide to theory and practice," Springer Science & Business Media, (1985)
- [6] B. Adamson and B. Lynch, "Language Program Evaluation: Theory and Practice," *Die Unterrichtspraxis / Teach. Ger.*, Cambridge University Press, (1999) DOI: 10.2307/3531893
- [7] S. Troudi and D. Nunan, "Research Methods in Language Learning," *TESOL Q., TESOL Quarterly*, (1995) DOI: 10.2307/3588081.
- [8] J. C. F. G. I. M. Jcgm, "Evaluation of measurement data — Guide to the expression of uncertainty in measurement," *Int. Organ. Stand. Geneva ISBN*, vol. 50, no.134, (2008) DOI: 10.1373/clinchem.2003.030528.
- [9] R. K. Yin, "Validity and generalization in future case study evaluations," *Evaluation*, vol. 19, no.3, pp.321-332, (2013) DOI: 10.1177/1356389013497081
- [10] D. T. Hickey, A. A., Ingram-Goble, and E. M. Jameson, "Designing assessments and assessing designs in virtual educational environments," *J. Sci. Educ. Technol.*, vol. 18, no.2, pp.187, (2009) DOI: 10.1007/s10956-008-9143-1
- [11] A. van Gelderen, R. Schoonen, R. D. Stoel, K. de Gloppe, and J. Hulstijn, "Development of Adolescent Reading Comprehension in Language 1 and Language 2: A Longitudinal Analysis of Constituent Components," *J. Educ. Psychol.*, vol. 99, no.3, pp.477, (2007) DOI: 10.1037/0022-0663.99.3.477
- [12] J. K. Phillips and H. H. Stern, "Fundamental Concepts of Language Teaching," *Mod. Lang. J.*, Vol. 70, No.1, (1986) DOI: 10.2307/328070
- [13] S. J. Savignon, "Communicative Language Teaching," *Theory Pract.*, vol. 26, no.4, pp.235-242, (1987) DOI: 10.1080/00405848709543281
- [14] T. O'Donoghue, "Standardized testing," in *Understanding Contemporary Education*, (2018)
- [15] J. Melitz, "English as a global language," in *The Palgrave Handbook of Economics and Language*, p. 583-615, (2016)
- [16] R. Kiely and P. Rea-Dickins, *Program Evaluation in Language Education*. Springer, (2005)
- [17] D. McCormick, "What Teachers Need to Know about Language," *System*, vol. 2, no.32, pp.291-294, (2004) DOI: 10.1016/j.system.2004.02.006
- [18] Z. I. Amankulova and S. K. Seisembieva, "Teaching English as a foreign language," *Analele Univ. din Craiova, Ser. Filoz.*, vol. 54, no.5, pp.414-418, (2011) DOI: 10.2307/811234
- [19] C. L. Machado-da-Silva, "Qualitative research & evaluation methods," *Rev. Adm. Contemp.*, vol. 7, no.2, pp.219-219, (2003) DOI: 10.1590/s1415-65552003000200018
- [20] R. F. Hudson, H. B. Lane, and P. C. Pullen, "Reading Fluency Assessment and Instruction: What, Why, and How?," *Read. Teach.*, vol. 58, no.8, pp.702-714, (2005) DOI: 10.1598/rt.58.8.1
- [21] J. B. Tomblin, "Language Development: An Introduction," *Top. Lang. Disord.*, (1990) DOI: 10.1097/00011363-199011000-00012
- [22] N. Bennett, W. R. Borg, and M. D. Gall, "Educational Research: An Introduction," *Br. J. Educ. Stud.*, (1984) DOI: 10.2307/3121583
- [23] D. McFarlane, "Evaluating Training Programs: The Four Levels," *J. Appl. Manag. Entrep.*, vol. 11, no.4, pp.96, (2006)