Список используемой литературы:


8. Перельман Я.И. Занимательная астрономия. – СПб.: ЭКСМО. 2017. 224 с.


10. Язев С.А., Комарова Е.С. Уровень астрономических знаний в обществе // Земля и Вселенная. 2009. № 5. С. 74-83.

УДК 372.881.111.1
ГРНТИ 14.35.09

APPROACHES TO ASSESSING LANGUAGE SKILLS AT HIGHER EDUCATIONAL INSTITUTIONS

Makovskaya Liliya Germanovna
Senior lecturer, Global Education Department
Westminster International University in Tashkent
12 Istiqbol Street, Tashkent Uzbekistan 100047

ПОДХОДЫ К ОЦЕНКЕ ЯЗЫКОВЫХ НАВЫКОВ В ВЫСШИХ УЧЕБНЫХ ЗАВЕДЕНИЯХ

Маковская Лилия Германовна
Старший преподаватель кафедры глобального образования
Международный Вестминстерский Университет в г. Ташкент
DOI: 10.31618/ESU.2413-9335.2020.7.76.945

АБСТРАКТ

Language assessment is widely discussed by specialists in applied linguistics and higher education. A growing body of literature has investigated the selection of appropriate scoring scales to be used in different teaching contexts. Given the significance of assessment in higher educational institutions, the article considers main approaches to testing language skills. It is explained that in the norm-referenced approach, students’ scores are shown in the relationship to other students in the group, university, or country. In the criterion-referenced approach, learners’ skills are assessed against a set of specific criteria. The article discusses the scoring scales for language assessment. Specifically, the holistic marking is based on the lecturers’ overall impression of the language assignment. University teachers assess language tasks analytically when they address each criterion separately. The article provides several recommendations for language teachers and increases awareness about the importance of developing marking scales for ensuring quality assessment in the university.

АННОТАЦИЯ

Оценка языковых навыков широко обсуждается специалистами в области прикладной лингвистики и высшего образования. Все большее количество литературы исследует выбор соответствующих шкал оценки, которые могут быть использованы в различных учебных контекстах. Учитывая значимость оценки языковых навыков в высших учебных заведениях, в данной статье рассматриваются два основных подхода к тестированию языковых навыков. При подходе, основанном на нормах, баллы студентов сравниваются по отношению к другим студентам в группе, учебном заведении или стране. Напротив, в подходе, основанном на критериях, навыки учащихся оцениваются на основе определенных критериев. В статье также обсуждаются шкалы оценивания, которые должны быть разработаны для оценки языковых навыков. В частности, целостный подход к оцениванию основан на общем впечатлении преподавателей от устного или письменного задания. Преподаватели университета могут также оценивать задания аналитически, если будут рассматривать каждый критерий оценки по отдельности. Данная статья содержит ряд рекомендаций для преподавателей и обращает их внимание на важность разработки шкал для обеспечения качества обучения в высших учебных заведениях.

Key words: assessment, norm-referenced, criterion-referenced, scoring scales, holistic, analytic, criteria
**Introduction**

Language assessment has always been one of the most discussed and controversial issues in higher educational establishments. Depending on the standards set by the education authorities, the main requirements for scoring are followed by all institutions; however, the ways language performance is assessed might vary. According to Nicholls, university teachers should always be “consistent, systematic, and constructive in the marking of students’ work” [13, p.115]. Therefore, it is of vital importance to implement a certain approach to assessment and develop criteria to be followed within a higher educational institution. As de Chazal highlights, it is significant for the university teachers to identify the way the assessment tasks are “marked and the scores are interpreted” [9, p. 300]. This will ensure the consistency, reliability and quality of teaching and learning process in the university settings.

Given the significance of assessment at the tertiary level, the article discusses two main approaches to assessment at university, i.e. norm-referenced and criterion-referenced measurements, as well as focuses on the holistic and analytic marking scales for assessing students’ productive language skills.

**Norm-referenced assessment**

In measuring language performance at the universities, two approaches are identified. The first one, norm-referenced assessment is defined by Bruce as the assessment, in which students’ scores are seen “in relation to the number of other people who received the same score” [4, p.201]. Alexander, Argent and Spencer clarify that scores (marks) are “expressed against a statistical average, or norm, for all the students” who perform the task [1, p.311]. For example, if a candidate has a score of 85 percent, then his/her score is higher than 85 percent of the total number of candidates, but lower than 15 percent of the rest [3, p.7; 9, p.301]. That is, the scores are ranked from highest to lowest and candidates are informed about the number (score) they achieved, but their performance is not interpreted in terms of any criteria.

On the global level, Brown exemplifies standardised tests, such as the Scholastic Aptitude Test (SAT) and the Test of English as a Foreign Language (TOEFL), which are taken by a large number of people. The author explains that such norm-referenced tests have “fixed predetermined responses” and are marked quite fast [3, p.7]. A similar instance is the entrance exams to the higher educational institutions, which use multiple-choice questions to identify the eligibility of the candidates for admission. To illustrate, one of the local universities (e.g. U1) accepts overall twelve prospective students, whereas the other university (e.g. U2) accepts overall ninety prospective students for the master’s degree studies in linguistics every academic year. These are the quota of these two universities, and those who achieve the highest scores (first 12 candidates for U1 and first 90 candidates for U2) in the entrance exams will be admitted. Thus, higher educational institutions set their own norm(s) that should be followed and usually no changes can be made because all the requirements are prescribed by the education authorities.

On the university level, according to Reece and Walker, norm-referenced assessment forms are traditional “end examinations and practical tests”, which are used to “ensure that standards are maintained” [14, p.417]. The exit tests administered at all the higher educational establishments provide certain norms to be followed by all the final-year students. To be eligible for the certificate/diploma, all the learners should meet these standards. In addition, Cohen et al. exemplify a national test of reading ability, and clarify that if 100 score is average, so when a learner achieves 120, then s/he is considered above average [6, p.398]. Such kinds of tests might be applied for the purpose of monitoring the quality of education at the higher educational institution, and sometimes the tasks of a similar format are completed by the students around the country. The results of this testing do not identify the level of learners’ knowledge, but show their abilities to meet the required standards and if they perform better or worse than “hypothesical average learners” of this age or level group.

Although norm-referenced assessment is widely spread for achieving the global (e.g. university entrance or exit tests) and institutional purposes (e.g. monitoring), this approach is not recommended for reaching the classroom goals because it “merely ranks test-takers” [4, p. 201]. For the language performance to be assessed, seeing the results in accordance with the norm is not sufficient. University lecturers should be able to identify the development of all language skills separately against the specified level criteria. Therefore, Brown highlights that criterion-referenced testing should be applied in the language classroom [3, p.7]. A detailed explanation is provided in the following sections.

**Criterion-referenced assessment**

The second approach to language assessment is criterion-referenced. It is defined as the assessment, in which criteria are set and learners are measured “according to whether they reach the level of attainment” [13, p.110]. Reece and Walker clarify that in comparison to norm-referenced measurement, all the learners can achieve high grades if they meet the requirements set by the university, or all the learners might fail if none of them achieve the performance standards [14, p.417]. Thus, there are no norms to be achieved as learners’ knowledge and skills are measured against the specific criteria and they get the score they achieve.

It is important to note that such kind of an approach to testing might be appropriate for assessing both receptive skills (listening and reading) and productive language skills (speaking and writing). The receptive skills are usually measured objectively by the use of various tasks (e.g., summary completion, matching, true/false, multiple-choice, and/or short
answer questions), which means the overall mark/grade for the task depends on the number of correct answers. So, if there are 50 items in the test, students can get from 0 (minimum score) to 50 (maximum score) correct answers depending on how well their knowledge and skills are developed. University lecturers can then report these results in different ways, e.g. in a numerical scale (e.g. 1-5 or 0-100%) or descriptive categories (e.g. from basic to proficient level). Table 1 shows a possible distribution.

**Table 1**

<table>
<thead>
<tr>
<th>Scores</th>
<th>Descriptor</th>
<th>Mark</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–10</td>
<td>failing</td>
<td>1</td>
<td>0–39</td>
</tr>
<tr>
<td>11–20</td>
<td>unsatisfactory</td>
<td>2</td>
<td>40–55</td>
</tr>
<tr>
<td>21–30</td>
<td>average</td>
<td>3</td>
<td>56–70</td>
</tr>
<tr>
<td>31–40</td>
<td>good</td>
<td>4</td>
<td>71–85</td>
</tr>
<tr>
<td>41–50</td>
<td>excellent</td>
<td>5</td>
<td>86–100</td>
</tr>
</tbody>
</table>

In comparison to receptive skills, the criteria for measuring productive skills are usually set and described separately for each skill (see the next article sections for further explanation). The number of criteria to be met by the students depends on the assessment task and university requirements. As Alexander, Argent and Spencer argue, criteria should be “specified in advance” because they provide a “transparent basis for grading performance” and ensure reliability among lecturers [1, p.312]. In addition, criterion-referenced measurement is recommended for assessing productive skills because it gives “developmental feedback” and is more helpful for the learners [4, p.201]. When students realise that they do not perform well in certain oral or written assignments, they have an opportunity to improve them for better performance.

Although norm-referenced and criterion-referenced measurements seem different, they might complement each other. For example, Cohen et al. explain that if lecturers use criterion-referenced tests, they can still compare the scores of students from different groups or institutions [6, p.399]. This allows identifying the performance level within the higher educational institution and across the region and/or the country (e.g. the highest and lowest results), but does not provide interpretation on the strengths and weaknesses of each student.

**Holistic scoring**

As criterion-based performance has become a key aspect in the language assessment, it has also become necessary to develop scoring scales to assess students’ language proficiency. Fulcher and Davidson explain that Liz Hamp-Lyons (1991) was among the first ones who distinguished different types of rating scales that can be used to assess performance in the second language context [10, p.249].

There are two main marking scales, i.e. holistic and analytic, that are widely used by the university language teachers all over the world. Weir clarifies that in holistic marking “an overall composite judgement” on the performance is made [15, p.181]. In addition, Biggs and Tang explain that university lecturers judge the assessment task by “understanding the whole in light of the parts” [2, p.214]. Thus, a teacher gives a mark/grade based on a number of specific criteria, which are developed for a speaking or writing task and grouped according to the performance level. Brown clarifies that “each point on a holistic scale is given a systematic set of descriptors” [3, p.242]. It is important to note that to ensure consistency in the description, each point should have the same number of criteria to be met and no additional components should be added. Table 2 shows the holistic scoring for assessing an oral presentation (i.e. assessment of speaking skills).

**Table 2**

<table>
<thead>
<tr>
<th>Mark</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>The presentation has an excellent logical flow of ideas. Fluent and confident speech; very effective eye contact and body language; skillful use of language.</td>
</tr>
<tr>
<td>4</td>
<td>There are mostly relevant ideas; appropriate flow of ideas. Some problems with eye contact and body language; several noticeable language errors that do not impede understanding.</td>
</tr>
<tr>
<td>3</td>
<td>The ideas are not arranged coherently. Lack of confidence and clarity in speech; body language might be inappropriate. Some language errors reduce effective communication.</td>
</tr>
<tr>
<td>2</td>
<td>The flow of ideas is occasionally impossible to follow. Poor eye contact and negative body language. Language errors seriously reduce effective communication.</td>
</tr>
<tr>
<td>1</td>
<td>Most parts of the presentation are missing; the flow of ideas is almost impossible to follow. Unclear speech; no eye contact and static body language. Language errors prevent communication.</td>
</tr>
</tbody>
</table>

As can be seen from the Table, the same criteria (e.g. organization of ideas, fluency, eye contact, body language, and use of language) are described for each mark/grade, but the description differs depending on the level of achievement. In case the higher educational institution has different marking schemes, e.g. percentage (0-100%) or letter grades (F-A), then the marks (1-5) should be converted accordingly.

There are a number of advantages in implementing holistic scales. For instance, Coombe, Folse, and Hubley explain that holistic marking is beneficial for teachers because it takes a shorter period of time to...
assess a large number of written papers [8, p.81]. In case the university lecturers assess more than a hundred of scripts or oral presentations, it will be much easier for them to give a score based on their overall impression of the written or oral production. Moreover, Brown believes that holistic scoring might guarantee “relatively high inter-rater reliability”, i.e. it ensures having no big discrepancy between marks/grades given by the teachers assessing the same production [3, p.242]. This is helpful for the teaching and learning process, especially if the department contains both novice and experienced language teachers. Another advantage of holistic scoring is that learners are not deprived of having a higher mark if one of the components is lower than others [8, p.81]. For example, if a student has an appropriate organisation of ideas and does not make many language mistakes, but is not confident enough and looks mostly at the assessors when presenting, his/her performance will mostly probably deserve ‘4’ mark rather than ‘3’ mark (see Table 2).

In spite of having several benefits, holistic scoring has a number of disadvantages. Jamieson argues that holistic score might not be helpful for formative assessment as it does “show students their strengths and weaknesses” [11, p.777]. For instance, if learners have ‘3’ mark for a writing task, they know what their overall performance is, but they do not realise what aspects have not been addressed. This also means that holistic scoring does not provide diagnostic information to the university lecturers [3, p.242]. Therefore, Katz believes that holistic tools are useful for marking students’ performance at the end of the academic year or for the placement purposes [12, p.329]. Another possible disadvantage of holistic scales is that teachers (especially novice or not well-trained) might have a tendency to either reduce or increase the overall mark by looking at the oral/written production as a whole [8, p.81]. Therefore, it is important for all the lecturers of the department to be aware of these drawbacks and be provided with professional training on assessment.

**Analytic scoring**

The second type of scoring scales is analytic. Weir explains that in analytic marking, “assessments are made in relation to each of a number of separate criteria” [15, p.181]. That is, each criterion/component is described and assessed separately. Katz highlights that it is important to use scoring guides because they “provide consistency in scoring as well as a clear picture of the criteria that will be used in judging a language performance” [12, p.329]. Table 3 illustrates the highest (5) and low (the lowest is 0) bands of the analytic writing rubric developed by the Cambridge English Assessment team and used for assessing writing performance at a B2 Level.

<table>
<thead>
<tr>
<th>B2</th>
<th>Content</th>
<th>Communicative achievement</th>
<th>Organisation</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>All content is relevant to the task. Target reader is fully informed.</td>
<td>Uses the conventions of the communicative task effectively to hold the target reader’s attention and communicate straightforward and complex ideas, as appropriate.</td>
<td>Text is well organised and coherent, using a variety of cohesive devices and organisational patterns to generally good effect.</td>
<td>Uses a range of vocabulary, including less common lexis, appropriately.</td>
</tr>
<tr>
<td></td>
<td>Irrelevances and misinterpretation of task may be present. Target reader is minimally informed.</td>
<td>Uses the conventions of the communicative task in generally appropriate ways to communicate straightforward ideas.</td>
<td>Text is connected and coherent, using basic linking words and a limited number of cohesive devices.</td>
<td>Uses everyday vocabulary generally appropriately, while occasionally overusing certain lexis.</td>
</tr>
</tbody>
</table>

Source: Cambridge Assessment English [5, p.2]

As can be seen from the Table, there are four criteria (i.e. content, communicative achievement, organisation, and language) that test-takers should meet for demonstrating their knowledge and skills at a B2 level in writing. Weir clarifies that in analytic marking “a level is recorded in respect of each criterion, and the final grade is a composite of these individual assessments” [15, p.189]. Thus, each criterion is described separately for each band (0-5), and therefore is assessed separately. For example, a student might receive ‘4’ for content, ‘5’ for communicative achievement, ‘3’ for organisation, and ‘4’ for language, which means ‘16’ will be the overall mark for the written task.

Certain oral and written tasks might require different distribution of scores, i.e. each criterion can
have specific weighting. In this case, apart from providing descriptors for each component (as given in Table 3), the percent for each criterion should also be specified. Table 4 demonstrates possible criteria weighting used for assessing oral presentation.

**Table 4**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Percent</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>20</td>
<td>0-4</td>
<td>5-8</td>
<td>9-12</td>
<td>13-16</td>
<td>17-20</td>
</tr>
<tr>
<td>Delivery</td>
<td>30</td>
<td>0-6</td>
<td>7-12</td>
<td>13-18</td>
<td>19-24</td>
<td>25-30</td>
</tr>
<tr>
<td>Language</td>
<td>30</td>
<td>0-6</td>
<td>7-12</td>
<td>13-18</td>
<td>19-24</td>
<td>25-30</td>
</tr>
<tr>
<td>Visual aids</td>
<td>20</td>
<td>0-4</td>
<td>5-8</td>
<td>9-12</td>
<td>13-16</td>
<td>17-20</td>
</tr>
<tr>
<td>Total score (max)</td>
<td>20</td>
<td>40</td>
<td>60</td>
<td>80</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

As shown in the Table, there are four main components (i.e. content, delivery, language, and visual aids) that have different weighting, either 20% or 30%. To illustrate, if a learner delivers a presentation, s/he can receive ‘16’ for the content, ‘18’ for the delivery, ‘21’ for the language, and ‘17’ for the quality of visual aids used during the presentation, which means that the overall score for the student’s performance is 72% out of 100% (which is the maximum). Coffin et al. highlight that although assessment criteria are given “specific weighting, markers need to exercise judgement” [6, p.79]. Thus, some training and practice might be required in the language department to ensure the quality assessment.

Higher educational institutions might have different grading systems (e.g. percent, mark, or letter grade), so the overall score given for the students’ oral or written production, can be converted to a mark or a letter grade. Table 5 shows one of the possible ways of such kind of mark/grade interpretation.

**Table 5**

<table>
<thead>
<tr>
<th>Conversion between percentage and letter grades</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fail</td>
</tr>
<tr>
<td>&gt; 45</td>
</tr>
<tr>
<td>D 46-50</td>
</tr>
<tr>
<td>C- 52</td>
</tr>
<tr>
<td>C 55</td>
</tr>
<tr>
<td>C+ 60</td>
</tr>
<tr>
<td>B- 65</td>
</tr>
<tr>
<td>B 68</td>
</tr>
<tr>
<td>B+ 70</td>
</tr>
<tr>
<td>A- 75</td>
</tr>
<tr>
<td>A 80</td>
</tr>
<tr>
<td>A+ 80+</td>
</tr>
</tbody>
</table>

Source: Biggs and Tang, Teaching for Quality Learning at University [2, p. 241]

Analytic scoring has a number of advantages. According to Katz, analytic tools are beneficial because they give “specific information about each component of a language performance” [12, p.330]. Therefore, teachers can identify what kind of language aspects should be paid more attention to and practiced in the classroom. In this case, university lecturers have an opportunity to develop activities and design lessons that might support learners in better understanding of the material. Coombe, Folse, and Hubley clarify that if the criteria are provided with explicit and detailed descriptors; they might be easily explained to the lecturers and applied in the assessment [8, p.83]. So, it is important to develop the criteria descriptors clear to all the teaching staff in the department, and if possible organise staff meetings for providing clarifications on the assessment criteria. In addition, Bruce explains that scoring analytically “provides more developmental feedback for students”, i.e. they understand what their strengths and weaknesses are [4, p.203]. For instance, a student might obtain a good score in the language use, but a low score in the logical development of ideas in writing; so s/he will realise that it is important for him/her to work harder on the improvement of linking the ideas in written tasks.

Although analytic scales are appropriate for language assessment, they might have several disadvantages. Coombe, Folse, and Hubley argue that analytic marking might be considered time consuming because lecturers should assess different aspects of the task [8, p. 83]. This means that analytic assessment will take longer in comparison to holistic marking. Moreover, some novice teachers might have difficulties in applying the analytic scales; therefore, additional training is required for them. Another drawback of this type of marking identified by Brown is a necessity to design different criteria and weighting for a variety of written and oral assignments [3, p. 246]. To illustrate, the use of visual aids might be an important component for some presentations and weigh up to 20%, and can be considered an additional, not required, component and not formally assessed in other oral tasks. Furthermore, sometimes university teachers might focus too much on the language use rather than other criteria to be met in the task. Thus, written production assessed analytically might be given lower marks than those assessed holistically [8, p.84]. To avoid this, understanding and proper use of the criteria should be ensured.

Given the importance of criterion-referenced assessment for the language performance, Coffin et al. believe that university lecturers should start developing the assessment criteria at the early stage of the teaching and learning process [6, p. 77]. Fulcher and Davidson explain that the way lecturers score “tasks needs to be considered as the tasks are being developed, not at some later stage” [10, p.257]. It is therefore significant to do it either at the end of the previous or the beginning of the current academic year, so that the criteria can be discussed first with the lecturers and then students in the classroom. Another significant point is the clarity of the criteria devised; thus, de Chazal explains that assessment criteria should “explicitly state what is being assessed” [9, p.302]. Such kind of transparency is important for university teachers and learners, as it makes the assessment reliable and fair for both parties.
Conclusion
To conclude, assessing language skills is an important component of the teaching and learning process at the tertiary level of education. In case the language tests are developed for the admission purposes, then norm-referenced assessment might be conducted. For achieving classroom goals, criterion-referenced assessment, which measures learners’ knowledge and skills against a set of specified criteria, is recommended. University lecturers should devise scoring scales, which allow measuring language tasks as a whole (i.e. holistically) or assessing every criterion provided for the oral or written assignment separately (i.e. analytically). Assessing students’ language skills in a proper way will ensure consistency in the teaching and learning process of the higher educational institutions and guarantee quality of education in the country.

References

STUDENTS ATTITUDES TOWARD ONLINE LEARNING OF CLINICAL SKILLS DURING COVID 19: CHALLENGES AND RESPONSES

Irma Manjavidze¹
MD, PhD, Full Professor,
Department of Clinical Skills and Multidisciplinary Simulation,
Tbilisi State Medical University, Tbilisi, Georgia

Dali Chitaishvili²
MD, PhD, Assistant Professor,
Department of Clinical Skills and Multidisciplinary Simulation,
Tbilisi State Medical University, Tbilisi, Georgia

Pirdara Nozdze³
MD, PhD, Associate Professor,
Department of Clinical Skills and Multidisciplinary Simulation,
Tbilisi State Medical University, Tbilisi, Georgia

DOI: 10.31618/ESU.2413-9335.2020.7.76.948

ABSTRACT
Covid 19 posed great challenges to the medical education system around the world. Because of restrictions the Department of Clinical Skills and Multidisciplinary Simulation of Tbilisi State Medical University (TSMU) had to start teaching Clinical Skills using Online Learning Format trying to maintain the basic principles and structure of Peyton’s 4-steps approach. The aim of our survey was to evaluate the students attitude to the Online Course of “Clinical Skills 3 “.

АНОТАЦИЯ
Covid 19 создал большие проблемы для системы медицинского образования во всем мире. Из-за ограничений Департамент Клинических Навыков и Мультидисциплинарной Симуляции Тбилисского Государственного Медицинского Университета (ТТМУ) пришлось перейти на онлайн-формат обучения, пытаясь сохранить основные принципы и структуру 4-степенчатого подхода Пейтона. Целью нашего исследования было оценка отношения студентов к онлайн-курсу "Клинические навыки 3".

1. Dali Chitaishvili
2. Irma Manjavidze
3. Pirdara Nozdze