CATEGORIES OF EVALUATION AND ASSESSMENTS

Usmanova Shokhsanam Avazovna

shoxsanam.shaxobova@gmail.com

English teacher at academic lyceum "International House Tashkent"

ESP teacher at MGIMO, Tashkent, Uzbekistan

Annotation: In education assessment establishes measurable and clear student learning outcomes for learning, providing a sufficient amount of learning opportunities to achieve these outcomes, implementing a systematic way of gathering, analyzing and interpreting evidence to determine how well student learning matches expectations, and using the collected information to inform improvement in student learning. The overarching aim of this research is to discuss some fundamental issues in second language assessment to provide classroom practitioners with knowledge to improve their test development skills. This research is to discuss some fundamental issues in second language assessment to provide classroom practitioners with knowledge to improve their test development skills.

Key words: assessment, evaluating, second language assessment, formative and summative evaluation, illuminative assessment, assessment categories.

Introduction

Educational assessment or educational evaluation[1] is the systematic process of documenting and using empirical data on the knowledge, skill, attitudes, and beliefs to refine programs and improve student learning.[2] Assessment data can be obtained from directly examining student work to assess the achievement of learning outcomes or can be based on data from which one can make inferences about learning.[3] As a continuous process, assessment establishes measurable and clear student learning outcomes for learning, providing a sufficient

---

[1] Some educators and education theorists use the terms assessment and evaluation to refer to the different concepts of testing during a learning process to improve it (for which the equally unambiguous terms formative assessment or *formative evaluation* are preferable) and of testing after completion of a learning process (for which the equally unambiguous terms summative assessment or *summative evaluation* are preferable), but they are in fact synonyms and do not intrinsically mean different things.

[2] Allen, M.J. (2004). *Assessing Academic Programs in Higher Education*. San Francisco: Jossey-Bass.

[3] Kuh, G.D.; Jankowski, N.; Ikenberry, S.O. (2014). *Knowing What Students Know and Can Do: The Current State of Learning Outcomes Assessment in U.S. Colleges and Universities*

amount of learning opportunities to achieve these outcomes, implementing a systematic way of gathering, analyzing and interpreting evidence to determine how well student learning matches expectations, and using the collected information to inform improvement in student learning.

Background knowledge

In discussing about language program evaluation, Richards (2001) presented three types of evaluation:

- Formative

- Illuminative

- summative evaluation.

Formative evaluation, as Richards pointed out, is utilized to find out the aspects of a program that are working well, not working well, and issues that need to be addressed. Some questions related to formative evaluation may involve seeking to find out if enough time has been spent on certain objectives or if the learning materials are well received. For classroom teachers, formative evaluation is an ongoing formal or informal evaluative process in which students are provided with various types of quizzes or tests which serve as a means for student learning.

Illuminative evaluation, according to Richards, is employed to find out how different aspects of a program are implemented and this type of evaluation is one way to seek to have "a deeper understanding of the process of teaching and learning that occur in the program, without necessarily seeking to change the course in any way as a result" (ibid, p. 289). According to Passerini and Granger (2000), "illuminative evaluations disclose important factors and issues emerging in a particular learning situation, factors which might have been overlooked by the instructor" (p.13). Examples of illuminative evaluation that Richards provided are finding out the strategies for error correction teachers employ or the strategies students use to deal with different text types. Illuminative evaluation for classroom teachers can be an instrument designed to assess a specific language point or problem to have a better understanding about students' difficulty in acquiring it, so that appropriate actions can be made.

Summative evaluation, the kind of evaluation most teachers and administrators are familiar with, is concerned with determining the effectiveness, efficiency, and to some extent the acceptability of a language program. Questions related to summative evaluation are if the course achieved its aims, what students learned, and if appropriate teaching methods were used. Summative evaluation can usually be final tests for classroom teachers.

Methodology

For classroom assessment, Mihai (2010) categorized it according to intention, purpose, interpretation, and administration. In regard to intention, an assessment can be informal when it is a spontaneous comment, or it can be formal when it is carried out in a systematic manner? In terms of purpose, an assessment can be formative if it focuses on the process of learning or it can be summative when it is used to measure student learning outcomes at the end of an education cycle. With respect to interpretation, an assessment may be used to compare students' performance with their peers' performance (norm-referenced) or it may be employed to compare students' performance with the course content (criterion-referenced).

Mihai (ibid) clarified that "whereas norm-referenced tests evaluate students in terms of their ranking to another, criterion referenced tests evaluate students in terms of their mastery of course content". The last category of assessment Mihai presented is administration which refers to the way an assessment is administered or delivered; an assessment may be classroom-based (small scale) when it is only used in the classroom or it can be delivered statewide or nationwide (large scale).

Assessment, moreover, can be conducted by either speaking or writing. Therefore, one more category of assessment may be added to those provided by Mihai: mode (oral or written). Table 1 provides a summary of types of assessment built upon the one provided by Mihai.

Category of Assessment	Type of Assessment
Mode	Oral
	Written
Intention	Informal
	Formal
Purpose	Formative
	Summative
Interpretation	Norm-referenced
	Criterion-referenced
Administration	Classroom-based
	Large scale

Table 1. The categories and types of assessment

Data collection and Analysis

There are 6 principles of second language assessment:

Validity

A test is considered valid when it reflects the test-takers' ability in a particular area and the test does not measure anything else. Validity is a complex concept in testing, but Brown and Abeywickrama (2010, p. 30) seemed to have well encapsulated the main attributes of validity.

They indicated that in order to achieve validity a test should:

- Measure only what it claims to measure,
- Not measure anything else,
- Rely as much as possible on empirical evidence,
- Involve performance that samples the test criterion,
- Offer meaningful and useful information about a test-taker's ability,
- Be supported by a theoretical rationale.

Reliability

A test is considered reliable if it is administered on different occasions and similar results are obtained. Brown and Abeywickrama (2010, p. 27) suggested the following ways to ensure that a test is reliable:

- It is consistent in its conditions across two or more administrations.
- It gives clear directions for scoring or evaluation.
- It has uniform rubrics for scoring or evaluation.
- It lends itself to consistent application of those rubrics by the rater.
- It contains items or tasks that are unambiguous to the test-takers.

Practicality

Practicality refers to the logistical, practical, and administrative issues involved in the process of constructing, administering, and rating an assessment instrument (Brown & Abeywickrama, 2010). Bachman and Palmer (1996, p. 36), on the other hand, defined practicality as "the relationship between the resources that will be required in the design, development, and use of the test and the resources that will be available for these activities."

Bachman and Palmer also added that practicality refers to the extent to which the demands of test specifications can be met within the limits of existing resources such as human resources

Equivalency and Authenticity

"An assessment has the property of equivalency if it is directly based on curriculum standards or instructional activities. Specifically, equivalency determines in what ways assessment design is influenced by teaching" (Mihai, 2010, p. 45). Equivalency is somewhat similar to authenticity which is defined as "the degree of correspondence of the characteristics of a given language test task to the features of a target language task" (Bachman & Palmer, 1996, p.23). Brown and Abeywickrama (2010) provided characteristics of a test that has authenticity as follows:

- It has language that is as natural as possible.
- It contains items that are contextualized rather than isolated.
- It includes topics that are meaningful, relevant, and interesting.
- It provides thematic organization to items, such as through a story line or an episode.
- It offers tasks similar to real-world tasks

Washback

Washback may have been called backwash, test impact, measurement-driven instruction, curriculum alignment, and test feedback (Brown & Hudson, 1998). Washback, according to Brown and Hudson, is the effect of testing and assessment on the language teaching curriculum that is related to it (p. 667). Likewise, washback is used to refer to the influence that a test has on teaching and learning. Washback, therefore, can both be beneficial and detrimental or positive and negative. Positive washback takes place when the tests measure the same kinds of materials and skills stated in the objectives and taught in the courses (Brown & Hudson, 1998). If a test encourages learning and teaching or if it provides opportunities for students and teachers to enhance the learning and teaching process, it is affecting language acquisition and instruction positively.

Nonetheless, if the test causes too much anxiety for the students, teachers and parents, the washback may be deemed as detrimental. Mismatches between the goals and objectives of the curriculum and tests can also be a source of negative washback. The example Brown and Hudson (ibid) gave is that "if a program sets a series of communicative performance objectives, but assesses the students at the end of the course with multiple-choice structure tests, a negative washback will probably begin to work against the students' being willing to cooperate in the curriculum and its objectives.

Result and Discussion

Students soon spread the word about such mismatches, and they will generally insist on studying whatever is on the tests and will ignore any curriculum that is not directly related to it. Moreover, other examples of negative washback, as discussed, are when the teacher teaches to the test or when students cram for the test. In fact, cramming for a test or teaching to the test does not truly promote enduring learning, so the main purpose of education may largely be ignored.

The literature on testing is extensive, and it is beyond the scope of this paper to include all the types of language assessment. This section reviews some commonly used assessment methods seemingly useful to second language teachers in general. A definition of each type of assessment will be provided, and the strengths and weaknesses of each type of assessment will also be examined. Readers interested in specific test types for individual skills and areas including listening, speaking, reading, writing, grammar, and vocabulary can find them in Brown. Coombe and Hubley (2003) noted that assessment in the field of English language teaching has come a long way from the days when it was simply equated with discrete-points, objective testing. They further added that although objective testing is still appropriate for certain purposes, assessment these days consists of a wide range of tools and techniques that range from testing an individual student's language ability to evaluating an entire program.

The method of assessment varies depending on the subjects and purposes of the assessment. Language testing is different from testing in other content areas as language teachers have more alternatives to make (Brown & Hudson, 1998).

Brown and Hudson (ibid) identified three basic assessment types:

a)   selected-response which includes true-false, matching, and multiple-choice assessments

b)   constructed response which includes fill-in, short answer, and performance assessments

c)   personal-response which includes at least conference portfolio, and self- and peer assessments.

Conclusion

It is clear that Assessment is an integral part of any education system and plays an important role in English as a foreign language context. Assessment establishes measurable and clear student learning outcomes for learning, providing a sufficient amount of learning opportunities to achieve these outcomes, implementing a systematic way of gathering, analyzing and interpreting evidence to determine how well student learning matches expectations, and using the collected information to inform improvement in student learning.

Used literature

1.   Allen, M.J. (2004). Assessing Academic Programs in Higher Education. San Francisco: Jossey-Bass.

2.   *Andrews, Fullilove, & Wong (2002). "Targeting washback—a case study". System.* **30** *(2): 207–223.*

3.   Bachman, L., & Palmer, A. S. (1996). Language testing in practice. New York, NY: Oxford University Press.

4.   Bachman, L., & Palmer, A. S. (1996). Language testing in practice. New York, NY: Oxford University Press.

5.   Brown, H. D., & Abeywickrama, P. (2010) Language assessment: principles and practices.White Plains, NY: Pearson.

6.   Brown, J. D., & Hudson, T. (1998). The alternatives in language assessment. TESOL Quarterly,32 (4), 653-675.

7.   Coombe, C. A., & Hubley, N. J. (2003). Assessment Practices. Alexandria, VA: TESOL.

8.   Kuh, G.D.; Jankowski, N.; Ikenberry, S.O. (2014). Knowing What Students Know and Can Do: The Current State of Learning Outcomes Assessment in U.S. Colleges and Universities

Yuldashova , N. A. qizi, & Ziyadulloyeva , M. S. qizi. (2024). ESSENCE OF DEVELOPING LEARNERS' COMMUNICATIVE COMPETENCE. GOLDEN BRAIN, 2(1), 572–575. Retrieved from https://researchedu.org/index.php/goldenbrain/article/view/6050