



*Approaching Validity of Online Testing under Covid-19 Pandemic
Teaching Protocol. A Case Study*

Hellalet Souhila

University Batna 2, (Algeria)

s.hellalet@univ-batna2.dz

Abstract ;

This paper attempts to investigate the issue of online tests, or more precisely, at home answering, validity under covid-19 pandemic and see the extent to which learners' answers reflect what their teachers have supplied them with and what they have really learned. The study was conducted at the English department at Batna2 university during the second semester of the 2019/2020 academic year and concerned twenty two teachers who used the online tests. An online questionnaire was sent to these teachers to gain quantitative and qualitative data, which enabled us approaching this topic in depth. The results obtained support the claim that remote testing conducted by our learners are at-home answering and do not reflect what and how online testing should be. Moreover, a general opinion was made that the actual way in which online tests are conducted in our department are far from ensuring validity. Other results are the basics of some suggestions made at the end of this study.

Article info

Received

18 December 2021

Accepted

26 January 2022

Keyword:

- ✓ Online Testing
- ✓ Assessment
- ✓ Validity
- ✓ Covid-19

1-Introduction

The last two years have marked a beginning of a new era in the social life in general and in the educational one in particular due to covid19. The end of which is still unknown .Teaching was not an exception and testing, as a consequence of teaching, was also affected by this change in both form and content. While best legislations are made by the ministry of higher education and put forward by the administration to keep students away from campus as much as possible to avoid this pandemic, the expectations were high, as learners follow online courses through a remote teaching /learning as a new brand. However, and supported by empirical evidence, these expectations did not come true as far as testing was concerned. Indeed, learners should be tested on what they have learned and their responses are supposed to reflect their real level, but online testing procedures imposed by the circumstances and exploited by our students' level of honesty renders the situation a more conflicting debate in the teacher's mind: whether to assume that the students answered online tests based on their true competences, or doubt the performances even if they were theirs; that was the question. The administration needs marks of students to fulfill the pass/fail selection, whereas the teachers still wonder if their tests answered at home and at ease were valid and, hopefully, reliable.

1.1 The Problem Statement

Covid-19 pandemic as an unexpected extraneous factor in the teaching /learning context urged teachers in the English department to resort to online testing as the

only way to assess learning. The circumstances of online test-taking and their newness to the teachers resulted in high scores obtained by the students compared to their real learning development. Next to our own experience of this issue and other teachers in a preliminary piloting study agreed that the validity of the tests, based on these scores, is questionable and the situation really needs an investigation.

1.2 The Research Objectives

The aim of this study is to gain some useful insights on how the online tests experienced for the first time in the English department could represent challenges to teachers and learners. In addition, the intention was to detect the deficiencies of these tests in ensuring valid assessment and to suggest some remedy to this issue.

1.3 The Research Questions

Following the above problem statement, two main questions are raised throughout our research:

- a- Were the tests constructed adequately with the criteria of online testing?
- b- Did the online test conditions ensure fairness among all the learners and reflect their real level?

1.4 The Research Hypothesis

Our research revolves around the following hypothesis:

The actual online testing conditions may not yield valid tests that will not be reliable to assess students' true acquired knowledge.

1.5 The Research Method

The present study falls onto action research. The nature of the topic and the information required calls for a description of the situation where tests are conducted online under inadequate circumstances. A descriptive approach will enable us to gain direct access to opinions and viewpoints held by our subjects being teachers practitioners of these tests. Along with the description facts, an analysis of the data collected adds substantial understanding of the possible remedy to the problem.

1.6 The Research Tools

A questionnaire of twenty items, as the main data gathering tool, was emailed to 22 teachers comprising our sample selected purposefully from an entire population of 68. In addition, informal interviews were held occasionally with teachers and some students whenever it was possible to discuss issues like test, scores, conditions and the like.

2. Literature Review

2.1 Introduction

The following chapter of this paper presents some fundamental issues related to the theory of language testing in its general framework. The aim is to gain insights on how language test specialists view testing and assessment as crucial steps in the teaching and learning process and how does testing mirrors through a useful feedback the adequacy of both instruction and evaluation. Commonly used types of language tests will be described and their relevance to teaching English as a foreign

language, which is the concern of this article, will be discussed.

2.2 Testing Vs Assessment

It is generally recognized by language teachers that the end of the teaching operation is summed up by an evaluative measure to see how well the instruction was accurate and learning was adequate. To do that, a variety of language tests are available each of which can be employed in particular setting and for specific purposes. While for academic purposes, language tests can assess a student's actual abilities or progress for placement reasons to know where he fits better, in professional settings, however, language tests decide whether a candidate has the required language skills needed for a specific task. According to LTI (Language Testing International), whatever the context, language testing can effectively measure a person's language abilities. Indeed, knowing how successful a person will be in a given task requires gathering information on his abilities to enable some predictions. As already stated by Norris (2000), cited in Saragih (2016), language tests are instruments or procedures for gathering particular kinds of information, typically information having to do with students' language abilities. Still according to him, teachers may use a variety of tests since language tests vary in formats, lengths, item types, scoring criteria, and media and we may differentiate among language test types according to such characteristics and the information provided by each. Whatever type is used, language tests assess aspects of a person's ability to understand or communicate in a particular

language. Perhaps, the difficulty arises when deciding on which type of test to be used and for what type of information required. Although each test is unique by its characteristics, but choosing one to be the fittest in a given area seems not easy for language teachers. Norris (2000), who has thoroughly examined this issue, concludes that selecting one test type over another based only on test's characteristics is like choosing between a hammer, a shovel or a screwdriver. What poses problems here is whether the choice was appropriate.

One way out of this dilemma is to decide whether the learners will be tested or assessed, and what abilities are sought to be known. Much debate over this idea took place, and a consensus was made that language teachers do not test and assess for the same reasons. If tests are tools used to gather information about a student's specific abilities, employing these information means assessing him. In clear terms, researchers say that making language assessment requires using language tests to accomplish particular jobs in language classrooms and programs. Describing how testing differs from assessing, Norris believes that practically, in language assessment, we first gather information in a systematic way with the help of language testing tools. In a concrete classroom setting, a teacher may decide about how well a student will be in a communicative exchange with native speakers based on his scores in an oral test. Likewise, a test of phonetic transcription may help the teacher assess the student's pronunciation skills. On the basis of this, decisions can be made on whether or not more instruction is needed.

In TEFL, a test is a measure used to know the level of skill or knowledge that has been reached by learners. It also means an evaluative procedure where examinee's performance in a specific domain to obtain scores that will be used to assess the amount of the acquired language knowledge using a standardized process. The test is used after the instruction is complete. Assessment refers to the process of gathering information using empirical data (test scores) to make decisions like altering instruction, planning reinforcement, or adjusting the teaching learning outcomes. On the basis of a test, the teachers may assess the learner's knowledge, skills and attitudes to further allow him a fair evaluation.

2.3 Formative Vs Summative Online Assessment

Modern technology has recently gained enough space in foreign language education and language teachers exploit such an opportunity in their classes not only for teaching but for assessing as well. Gikandi et al. (2011), who have investigated how online assessment has become widely employed in higher education, assert that its benefits may serve teachers in many constructive ways to boost the teaching and learning process. Although these researchers have focused more on formative assessment and its utility in online testing, they provided a comprehensive analogy between the summative and the formative types. Authors like Hargreaves (2008) and Arnold (2010) have also pointed to the fact that while assessing students language knowledge, a distinction is made between summative assessment which measures

what students have learned at the end of an instructional unit, end of a course, or after some defined period, and formative assessment which is commonly applied in the classroom as a source of ongoing feedback with the aim to improve teaching and learning (Hargreaves, 2008 in Gikandi et al. (2011) .

Saragih (2016), who holds a similar opinion on this issue, identifies two main functions for assessment. To her, assessment is formative when the learners are evaluated in the process of forming their competencies and skills with the aim of helping them carry on that growth. Summative assessment, however, does not necessarily target future progress but simply indicates what learners have indeed grasped in the whole teaching term. To researchers like Challis (2005), summative assessment can also refer to determining that the targeted objectives of learning have been met or that the needed levels of competence are actually attained. Formative assessment, which is referred to as assessment for learning (AFL), is an ongoing process of gathering information during the course of instruction with the aim to support learning (Oosterhof et al., 2008, pp. 76–77; Vonderwell et al., 2007). Recent research has expanded the difference between these two types of assessment and provided a simple and practical framework for its face to face as well as online uses: assessment for learning (AFL) in formative assessment, as opposed to assessment of learning (AoL) in summative assessment, has arisen as a novel term in recent debates on tests (Fitriyah & Jannah 2021). Further, they added that, and in accordance with the

ongoing changes in learning theories, assessment in general has shifted from AOL to AFL to enable evaluators move from evaluation for qualification purposes to an evaluation where learners carry out themselves and peer assessment and build critical knowledge through formative awareness. Lee (2007), in the words of these authors, has summarized the whole point in an exclusive comprehensible picture: Learning is a goal in the AFL process, and assessment is a tool for achieving that goal.

Research has also shown that online assessment regardless of its summative or formative natures has many advantages to teachers and learners although few drawbacks can be remarked. In the Indonesian Journal of English Language Teaching and Applied Linguistics IJELTAL (2021), one can read that online assessment is no longer a teacher's free choice of a testing procedure, but rather a compulsory medium of carrying out an evaluative measure during the last two years. Covid-19 pandemic urged testing professionals resort to online facilities as safe measures for assessment instead of engaging in face to face or paper-based test with all the potential risks they may generate. Still according to this same source, online assessment can be useful in the teaching learning operation in many sorts of ways.

Spivey and McMillan (2014), cited in IJELTAL (2021), concluded that compared to offline assessment, online testing has many advantages:

- It allows for adjustable testing cycles in which the test can be conducted,

- Online software can be set up to automatically randomize the order of questions as well as the set of responses for multiple choice and matching questions.
- Different levels of feedback, such as a test score, a test score with accurate answers, or a test score with detailed solutions, may be provided,
- It is possible to set up preferred online testing systems to include clues or prompts as to where the text or course notes will find support to answer a question.

In general, the benefits of online assessment are concrete and better exploited when they are approached through a variety of measures. Varying and multiplying the testing tools leaves few chances for undesirable effects to occur. Gikandi et al. (2011) say that variety of approaches facilitate a multidimensional perspective to online assessment. To them, a synchronous discussion forums, self-test quiz tools, and e-portfolios are examples whose effectiveness depends on innovative and appropriate utilization in order to make online formative assessment an effective pedagogical strategy.

2.4 Validity in Online Assessment

Validity is an integral component in language testing, and researchers emphasise its paramount role in making a good test. Pioneers in psychology and education have all pointed at the crucial and critical nature of validity. Lado (1961) for instance, remarked that, although a construct, validity ensures that a test measures what it is meant to measure. Contemporary researchers stand on a similar position and some of them attribute to validity an exclusive function. A test, according to Brown & Bailey (2008), is

valid if it measures what it has to measure and nothing more (Giraldo, 2020). Perceiving validity as an abstract concept makes its measurement complex even for professionals in the field. To render it practical, Messick (1989) links validity to the inference that is made from the test results. Later, Shaw and Crisp (2011) expanded this view and argued that validity is approached based on « the appropriateness of the inferences and uses made from assessment outcomes, including some considerations of the consequences of test score use” (p.14). Gikandi et al (2020), who quoted these authors concluded that validity emphasises more what we infer from the test results and what decisions we make accordingly than applying it for the test itself.

Testing professionals suggest many types of validity and relate each type to a specific testing purpose. Although no ideally classification of the different types of validity was advocated as the fittest, but researches converge to identify three main criteria commonly observed in approaching validity in language classes :

2.4.1 Aligement of Assessment with Learning Outcomes

According to Bergeron (2020), cited in the American Association of Colleges and Universities, alignment is one of the most important considerations when addressing the validity of the assessment results. In theory, this is what teachers endeavour to achieve. In practice, however, this task seems rather vague demanding well planned strategies. To demystify this issue for teachers, the Teaching and Learning Transformation Centre (TLTC University of Maryland) suggest a guideline to make

sure the assessment has met the learning goals advising teachers to:

1. Formulate a clear and succinct learning goal (or goals) for students.
2. Decide what students should be able to do if they have met those learning goals.
3. Develop an appropriate assessment instrument (a test, essay, project, etc.) and a scoring rubric.
4. Assess students' mastery of the learning goals given their performance on the assessment instrument.
5. Reflect why students did or did not master the learning goals, and develop strategies to help them be as or more successful in the future.

2.4.2 Adjustment of the Test Construct

Adjusting the test construct aims at mapping the test items with the quality of the knowledge expected from the learners. For attaining such an objective, the Center for Innovative Teaching and Learning (ITLC University of Indiana) simplifies the process particularly if language tests require students writing essays:

- a- Specifying whether we assess synthesising or analyzing.
 - b- Ask questions which require students to discuss, compare, criticize, and so on.
 - c- Design questions that prompt student's organisation of the answer.
 - d- Write items that define the parameters of expected answers as clearly as possible.
- Teachers, who adjust test constructs following such a framework, will be able to assess a large scope of the potential learning rather than focusing on discrete point testing. Adjusted test like these target more general mastery of the language.

2.4.3 Ensuring fairness

Fairness in assessment is another characteristic of a valid test. According to the International Bureau of Education (IBE), the term refers to "the consideration of learner's needs and characteristics, and any reasonable adjustments that need to be applied to take account of them". This means that learners think a test is fair when they are informed about, have enough knowledge and motivation to take it, and accept the testing process. In the same sense, Tierney (2016), considers that fairness in educational assessment is not easy to achieve and needs careful planning given its complex nature. Suskie (2018) agrees with that and conceives the idea of fairness in a very expressive picture. In her words: "A test question on quantitative skills that asks students to analyze football statistics might not be fair to women". Indeed, a test built on a biased assumption does not fairly assess all the students taking the test unless it uses constructs that are equally familiar to all of them. Again, approaching fairness in assessment, Tierney (2016) recommends opportunity to learn, a constructive environment, and evaluative thinking as three main components that enable conditions for a fairer assessment. This means that learners should have experienced equal chances for learning before test taking and that the conditions under which they are being tested are not significantly unequal.

3. Results and discussion

The participants' answers to our questionnaire are presented in the following section. The responses are interpreted through our inferences of facts

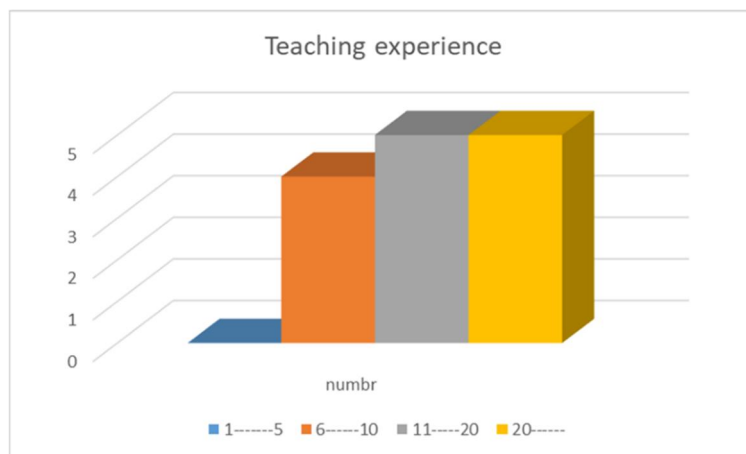
and supported by statistics displayed in tables and figures for illustration and clarification. Of course, we have assured our respondents that their contribution will be anonymous to encourage them say all what the items could not cover. Parts of answers were embedded in our interpretation while the most prominent ones are reproduced for the sake of emphasis.

Item 1: Teaching experience

The majority of our respondents share a long career in teaching beyond ten years of experience. This fact led us to assume right at the beginning of the data analysis that it

will act positively on their perception and conceptualization of online testing ensuring objective evaluation through valid online testing procedures. Of course, after much practice, teachers get acquainted with the strategies underlying the testing system. Such an experience may assist them designing test that ensure a correlation between students' efforts and their exam result as a way to enhance learning performances. Nevertheless, and as will be seen later, our respondents' answers to the coming items do not seem to support our claim.

Fig.1. Teaching Experience



Item2. How often did you use online testing before covid-19 pandemic?

Table 1. Teachers' frequency of using online testing

Always	Sometimes	Never
0	2	14

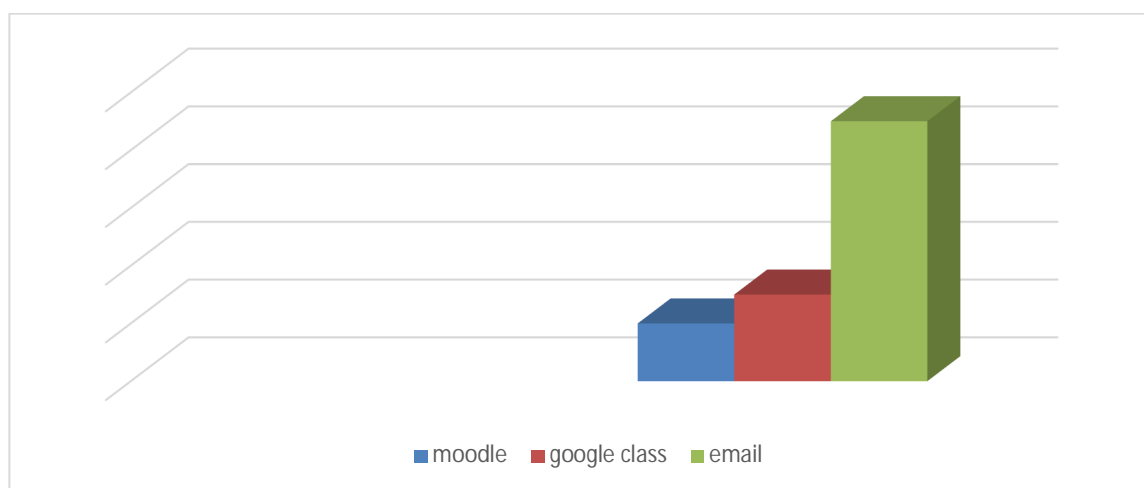
As table 1 shows, out of the seventeen who responded to this item, fourteen teachers have never used online testing

before the pandemic. To us, this is not in line with their responses to previous item which shows that their majority is

not novice, but well experienced teachers. Their considerable long experience did not assist them search for other innovations in test taking such as online techniques and to practice them at ease before they become an urgent need. Here, we still wonder: Is it

Item3: Which online testing platform you have used?

Figure 2. Online testing platforms



We have suggested a range of online testing platforms used for online testing by professionals. What should be noticed is that we have included email as platforms although it does not fit that pedagogical purpose for it is the less reliable among all others.

While asked on which online testing platforms are used, the great majority of the respondents confessed that they recourse to e-mailing as a tool. Knowing that teachers have no evidence of the students honesty while answering the test, this may reveal that one criteria of online testing (integrity) is not met here. Students may answer at home and at ease consulting references and

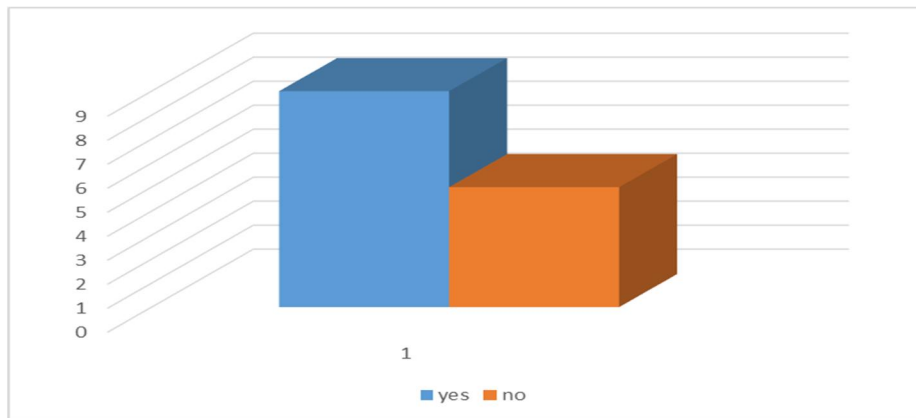
due to their unfamiliarity with this evaluation mode, or rather is it due to the impracticability of the online testing mechanisms in our department? Their responses to the next item helped us to answer these questions

presenting them as theirs, misleading the teacher in gathering the needed information for a summative assessment; which, in turn, will not enable a fair formative assessment. In our opinion, this is revealing of two elements:

First, despite the increasingly sophisticated modern technologies, online platforms did not yet gain attraction of our teachers. Rather, they prefer relying on a less valid testing tool than adventuring in a more secure but demanding one. Second, whether or not students answer by themselves at home, their performances do not truly reflect their language knowledge while answering an emailed test.

Item 4. When designing an online test, do you adjust its format to meet the requirements of theonline environment?

Figure 3. Adjustment of test formats



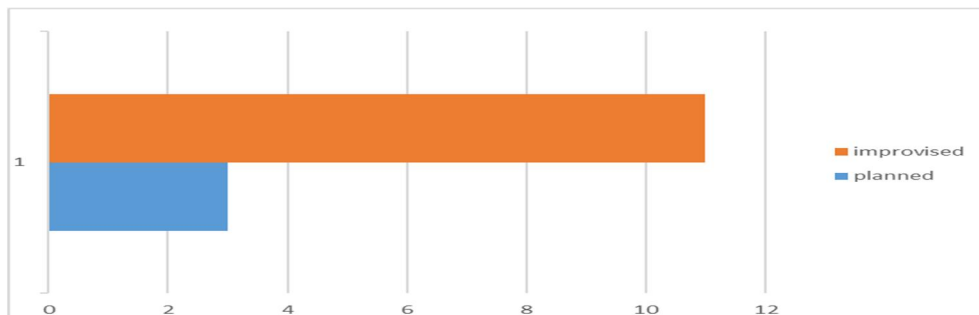
As shown in this figure, most of the teachers say they adjust the test format to fit the online testing environment and its requirements. Indeed, in theory, an online test is adjusted to meet some comfortable conditions for test taking ensuring low level of anxiety and stimulating students' involvement and motivation. In practice, however, this does not match with what goes on in our department. When asked to

name some criteria of online test validity, only two teachers have mentioned the most common ones. We may understand here that if a teacher ignores the characteristics of a valid test, how could he claim to adjust it for specific circumstances? Whether we believe the teachers' opinions, or trust statistics of their answers, the fact remains that the issue of online test validity in our department did not yet reveal all its secrets!

Item4. Was the adjustment of your format to meet the requirements of the online environment during the pandemics planned or improvised?

Our teachers' responses to this item are better illustrated in the following figure

Figure 5. Planning Vs Improvising



The intention behind asking this question was to know if really the information

provided by our respondents in the previous item can be crosschecked here.

Teachers who are willing to adjust their test formats to meet the requirements of a specific test environment (covid pandemic in our case) should plan to do that. This is another indication that most of the teachers used online tests to fit a situation rather than to meet the requirements of a valid and reliable testing measures. Online test research says we should adjust the online test format for specific situations with the ability of control over test taking

mechanisms, test objectives, and students real level. It is too doubtful to assume that these improvised adjustments fit such purposes. One of the teachers has overtly declared this: “ *...improvised to meet the students' level keeping in mind that in-class thorough explanations were lacking The uploaded worksheets alone can never substitute the teacher's presence*”.

Item 6. To what extent did the online testing results accurately correspond to the students' real level?

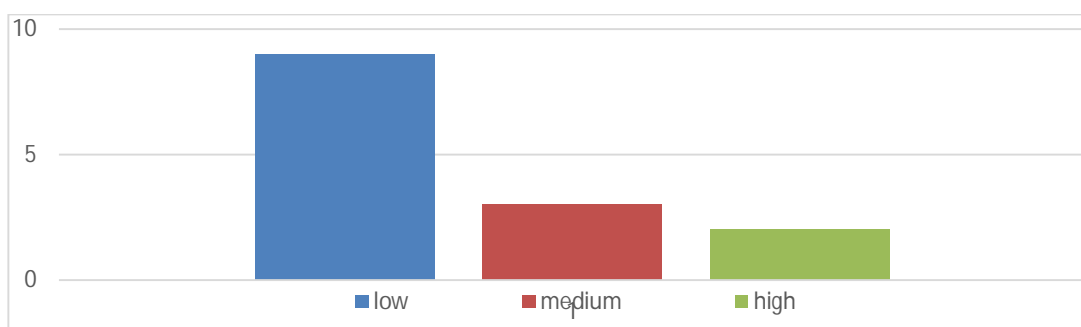
Table 4: correspondence between tests and levels

Far extent	Medium	Low
2	3	9

Out of the fourteen teachers who answered this question, the absolute majority do not trust their online tests to reflect the

learners' real level. To figure out this situation, the following diagram is more expressive.

Figure 6. teacher's acknowledgments



This item gauges the teachers' appreciations of their students' real learning levels by means of the designed online tests. It was not our intention to quantify teaching nor to doubt the teachers' skills in transmitting knowledge. Simply we were interested

to see how far the topic we are investigating concerns other teachers in our department and try to find through a collaborative work best remedy to the issue of online test validity. What is worth noting here is the fact that the teachers confess that the tests' results do

not reflect accurately their learner’s level. Astonishing enough was their acknowledgment in many cases that they have no other alternative at hand. A worth asking question arises here: Do **Item7. Did your online test target ?**

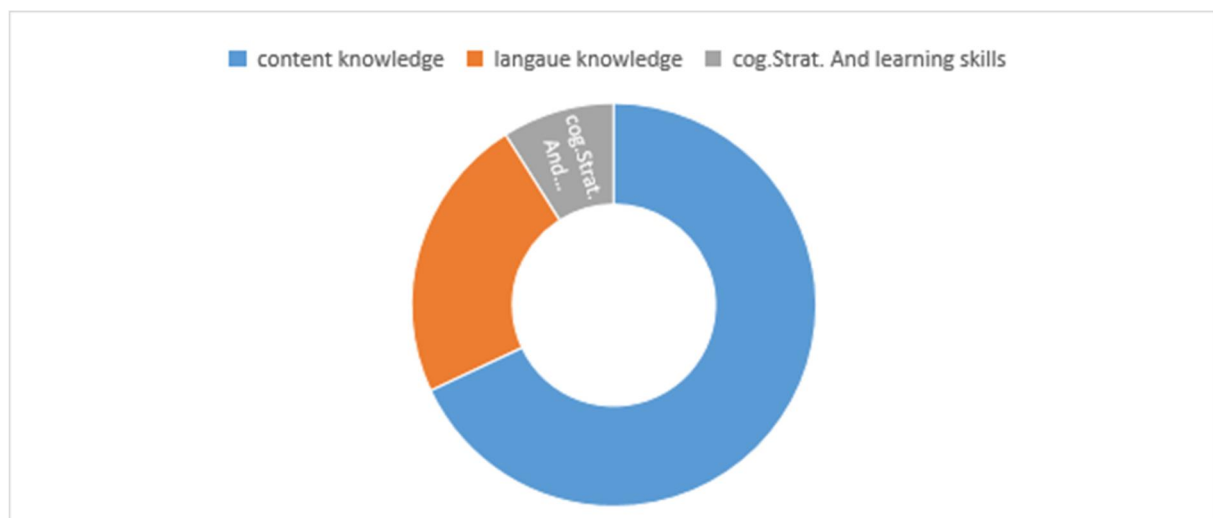
we really, as teachers, have to conform to certain regulations in spite of their inappropriateness in our teaching tasks? I believe the answer requires a whole investigation itself.

Target	Content knowledge	Language knowledge	Cognitive abilities and language skills
Nbe	13	05	02
%	68	23	09

Teachers’ answers to this item were summarized in this table. At the first glance, one can notice that teachers care more about course content when designing online tests. This can be understood if the assessment they opt for is formative, which seeks primarily to identify weakness in both teaching and learning more than evaluating learning

and achievement outcomes. Nevertheless, this should not be the case in our department: tests, whether conducted online or offline, are meant for quantifying learning to identify who passes and who does not. The proportions of the teachers’ focus in online tests targets are well represented in the following figure:

Figure 8. teachers’ focus in test targets



It is widely agreed among researchers that online tests are meant to check students’

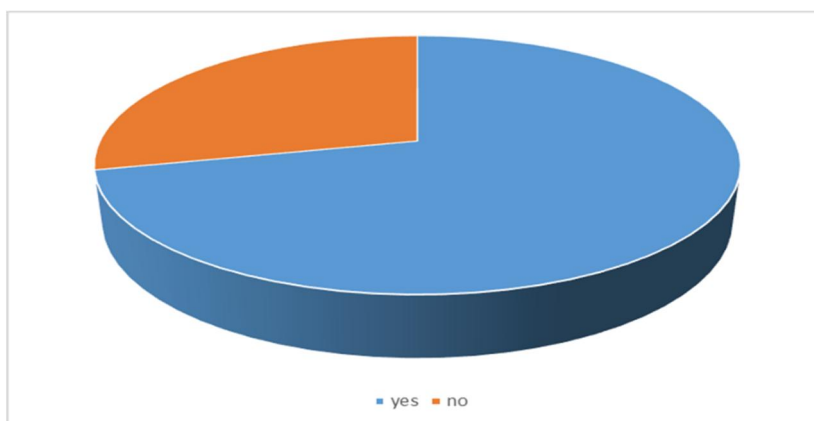
cognitive strategies and learning skills as these types of tests rely on higher- order

thinking. The test taker has to demonstrate a skill of joining rapidly, correctness, quick recall of information, and a required easiness of manipulation of knowledge to respond to the test items. But the responses to this item are far from such a perspective, since 70% of the online tests held at our department target content knowledge. Following the teachers's responses, "*Many factors have to be taken into consideration such as the tests' length (issue of technological devices' affordability and accessibility i.e. many students do not have computers and/or internet access so they go to the cybercafes), the documents' size (due to poor internet connection), the questions' clarity, etc*". Once more, reality teaches us what theory does not aim at.

Item 9. Did your online testing ensure alignment with the course content and learning outcomes?

This question was asked to know if test validity was approached at least partly through aligning the course content and learning outcomes. So far, the teachers in the English department have shown considerable concern with relating the course content to the learning outcomes, but not the test content with these two, often at the detriment of test validity. What supports our claim here are their responses to this item where most of them affirmed so as clearly shown in this diagram.

Figure 9. Alignment with course content and learning outcomes



Based on the previous responses, and consulting theory, tests held online in our department cannot ensure alignment with course content and learning outcomes. Simply, because they do not land themselves on basic criteria of validity. The teachers' answers here suggest the

opposite. If the great majority of them think their online tests yield such an alignment, it can partly be due to insufficient familiarity with that mode of testing. Added to this, is the total absence of control means over students' answers settings which favor low

level of integrity and high rate of dishonesty.

3.1 Recommendations

Our investigation of this topic was brought to its conclusion with the conviction that the potential of engaging both teachers and students in the online assessment methodology is not well exploited. There is still much to be done, and more future research has to take place since online education seems gaining more space over traditional pedagogy. We believe our contribution may assist further actions by considering what is recommended in the following points:

- In-service teacher training is needed as the summative assessment rests upon valid online testing in specific teaching courses.
- Technological softwares and online testing platforms should be widely introduced and accessible to both testers and testees.
- Online tests should target high-order thinking and learning skills to prevent temptations of cheating and thought stealing.
- When answering an online test, time should be reduced to the minimum.
- Teachers should favour the assessment formats which encourage learners express their own thoughts.
- Higher order thinking should be more addressed while taking an online tests. Students will have the chance to be assessed on their deeper levels of processing available knowledge.
- The scoring systems should be designed in a way to quantify the true constructed learning outcomes.
- Online test questions should ensure scientific integrity thorough requiring

students to demonstrate skills in synthesising, analysing, and arguing in academic styles of writing.

4. CONCLUSION

This small-scale survey was launched to explore the possibility of gaining profit of modern technology and applying it to the service of teaching generally and testing particularly. It tackled the question of whether and to what extent online testing improvised during the pandemic as a magic evaluation tool to complete the assessment task and end safely the whole academic year. Good intentions were behind such an innovation but total disparity added to a low level of integrity, conjugated with a high level of difficulty in practicing this testing modality reduced the best expectations to a very hard reality. The majority of the teachers have a considerable experience but a few of them have already experienced online testing before the pandemic era. Even for these ones, the issue of validity is an era where much research and practice are still needed. Despite the good will of teachers to construct their online tests adequately to the learning outcomes, the testing environment in the English department does not actually enable them attain such a target. The present conditions under which the online tests are conducted fit the authorities and the administration more than teachers and learners because they enable only the move-on policy more than the learning adequacy. If these tests are really meant to fulfill what they should do, then lot of measures should take place. It is fair to understand that the teachers improvised online tests as the best solution at hand to end up their teaching tasks, but

theory recommends that tests held online should be planned to align with the course content and the learning outcomes. Without such connections, little benefit from these tests will be expected. Whether students are given open-book test or close-book test formats, the fact remains that the degree of their integrity while answering the test is doubtful. Really hard to conceive, but taking the test at home and at ease incites them for plagiarism and dishonesty. It was not surprising to see that almost all the teachers used the e-mail as a platform for online test taking. What was rather astonishing was their confession that scores obtained through these tests do not reflect real learning. This is what goes in the direction of our hypothesis that addressing validity in such testing ways is still hard to achieve.

5. Bibliography List:

- Arnold, S.D. (2010). Student e-cooperative: Computer mediated group projects. In the International Technology, Education, and Development Conference Proceedings CD (ISBN: 978-84-613-5538-9). Valencia, Spain.
- Brown, J. D., & Bailey, K. M. (2008). Language testing courses: What are they in 2007? *Language Testing*. 25(3), 349–383.
- Challis, D. (2005). Committing to quality learning through adaptive online assessment. *Assessment and Evaluation in Higher Education*. 30(5), 519–527.
- Gikandi J.W. et al. (2011). Online formative assessment in higher education: A review of the Literature. *Computers & Education*. 57 (2011) 2333–2351
doi:10.1016/j.compedu.2011.06.004
- Giraldo, F., D. (2020). <https://revistas.udistrital.edu.co/index.php/calj/article/view/15998>.
- Hargreaves, E. (2008). *Assessment*. In G. McCulloch, & D. Crook (Eds.), *The Routledge international encyclopedia of education* (pp. 37–38). New York: Routledge.
- Fitriyah, I., Jannah, M. (2021). Online Assessment Effect in EFL Classroom: An Investigation on Students and Teachers' Perceptions. *IJELTAL*. 5(2), 265-284 DOI: <http://dx.doi.org/10.21093/ijeltal.v5i2.709>
International Bureau of Education (IBE) from <http://www.ibe.unesco.org/en/glossary-curriculum-terminology/f/fairness-assessment>
- Lado(1961)
ITLC, Test Construction University of Indiana from <https://citl.indiana.edu/teaching-resources/assessing-student-learning/test-construction/index.html>
- Lee, I. (2007). Assessment for learning: integrating assessment, teaching, and learning in the ESL/EFL writing classroom. *Canadian Modern Language Review*. 64(1), 199–213.
- Messick, S. (1989). Validity In R. Linn (Ed.), *Educational measurement* (pp. 13–103). New York: Macmillan.
- Norris.J.M. (2000). Purposeful Language Assessment: Selecting the Right Alternative Test. *English Teaching Forum*. Number 3
- Saragih, F. H. (2016). Testing and Assessment in English Language Instruction. Retrieved from <https://media.neliti.com/media/publications/74656-EN-testing-and-assessment-in-english-langua.pdf>
- Shaw, S., & Crisp, V. (2011). Tracing the evolution of validity in educational measurement: Past issues and contemporary challenges. *Research matters*. Retrieved 16 May, 2011, from. A Cambridge Assessment Publication.
- Spivey, M. F., & McMillan, J. J. (2014). Classroom versus online assessment. *Journal of Education for Business*. 89(8), 450–456. <https://doi.org/10.1080/08832323.2014.937676>

- Suskie.L. (2018). 6 Ways to Ensure your Assessment Practices are Fair and Unbiased
<https://www.wiley.com/network/featured-content/6-ways-to-ensure-your-assessment-practices-are-fair-and-unbiased>
- Tierney, R. D. (2016). Fairness in educational assessment. In M. A. Peters (Ed.), *Encyclopedia Of Educational Philosophy and Theory*. Singapore: Springer Science+Business Media. DOI 10.1007/978-981-287-532-7_400-1)
- Oosterhof, A. Conrad, R. M., Donald, P. (2008). *Assessing learners online*. New Jersey: Pearson. pp. 76–77;
- Vonderwell, S. & Liang, X., Alderman, K. (2007). Asynchronous discussions and assessment in online learning. *Journal of Research on Technology in Education*. 39(3), 309–328