ing, teaching practices and teaching content and lead to the shaping and renovation of education systems and teaching programs.

References

Al-Kharusi, H., Aldhafri, S., Alnabhani, H., & Alkalbani, M. (2012). Educational assessment attitudes, competence, knowledge, and practices: An exploratory study of Muscat teachers in the Sultanate of Oman. *Journal of Education and Learning*, 1(2).

Al-Nouh, N., Taqi, H. & Abdul-Kareem, M. (2014). EFL Primary School Teachers' Attitudes, Knowledge and Skills in Alternative Assessment. *International Education Studies*, 7(5), 68-84.

Birgin, O., & Baki, A. (2009). An investigation of primary school teachers' proficiency perceptions about measurement and assessment methods: The case of Turkey. *Procedia Social and Behavioral Sciences*, 1, 681-685. http://dx.doi.org/10.1016/j.sbspro.2009.01.119

Cirit, N. C. (2015). Assessing ELT Pre-Service Teachers via Web 2.0 Tools: Perceptions toward Traditional, Online and Alternative Assessment. *The Turkish Online Journal of Educational Technology*, 14(3), 9-19.

Cohen, L., & Manion, L. (1994). *Research Methods in Education*. 4th eds. London: Routledge.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in Education.* London and New York: Routledge Falmer.

Garrison, C., & Ehringhaus, M. (2007). Formative and summative assessments in the classroom. Westerville, OH: Association for Middle Level Education. https://www.amle.org/portals/0/pdf/articles/Formative_Assessment_Article_Aug2013.pdf

Ghazali, N., Yaakub, B., & Mustam, A. (2012). "Why do we need to change?" Teachers' attitude towards school-based assessment system. SCR London's First International Conference on Social science and Humanities in the Islamic World (28-30 May 2012).

Gonzales, R., & Aliponga, T. (2012). Classroom assessment preferences of Japanese language teachers in the Philippines and English language teachers in Japan. *MEXTESOL Journal,* 36(1).

Guerin, E. (2010). *Initial findings from a pilot Italian study of EL teachers' stated language assessment knowledge-base and needs*. Papers from the Lancaster University Postgraduate Conference in Linguistics & Language Teaching, Vol. 4: Papers from LAEL PG.

Harlen, W., & James, M. (1997). Assessment and learning: Differences and relationships between formative and summative assessment. *Assessment in Education: Principles, Policy & Practice*, 4(3), 365–379.

Kiomrs, R., Abdolmehdi, R., & Naser, R. (2011). On the interaction of test washback and teacher assessment literacy: The case of Iranian EFL secondary school teachers. *English Language Teaching*, 4(1), 156-161.

Knight, P., & Yorke, M. (2003). *Assessment, learning and employability*. Maidenhead: Society for Research into Higher Education & Open University Press.

Metin, M. (2011). The examinations of teachers' attitude towards performance assessment with respect to the different variables. Energy Education Science and Technology Part B: Social and Educational Studies, 3(3), 269-284.

Muhr, Thomas, and Susanne Freise. 2004. User's Manual for Atlas.ti 5.0 (2nd Edition). Germany: Scientific Software Development.

O'Leary, M. (2006). Towards a balanced assessment system for Irish primary and secondary schools. *Oideas*, 52.

Papakammenou, I. (2016). A Washback Study of the Teaching Practices Used in EFL Exam Preparation Classes in Greece. In D. Tsagari (Eds.), *Classroom-based Assessment in L2 Contexts*, (pp.118-137). Newcastle upon Tyne: Cambridge Scholars Publishing.

States, J., Detrich, R. & Keyworth, R. (2018). Overview of Summative Assessment. Oakland, CA: The Wing Institute. https://www.winginstitute.org/assessment-summative.

Stiggins, R. (2001). The unfulfilled promise of classroom assessment. *Educational measurement. Issues and Practice*, 20(3), 5-15. http://dx.doi.org/10.1111/j.1745-3992.2001.tb00065.x

Tangdhanakanond, K., & Wongwanich, S. (2012). Teacher attitude and needs assessment concerning the use of student portfolio assessment in Thailand's educational reform process. *International Journal of Psychology*, 10, 7-8.

Tante, A. (2010). The purpose of English language teacher assessment in the English-speaking primary school in Cameroon. *ELTED*, 13.

Tsagari, D. (2009). Revisiting the concept of test washback investigating FCE in Greek language schools. *Research Notes,* 35, 5-10.

Vogt, K., & Tsagari, D. (2014). Assessment Literacy of Foreign Language Teachers: Findings of a European Study. *Language Assessment Quarterly,* 11, 374–40

Watt, H. (2005). Attitudes to the use of alternative assessment methods in Mathematics: A study with secondary mathematics teachers in Sydney, Australia. *Educational Studies in Mathematics*, 58, 21-44. http://dx.doi.org/10.1007/s10649-005-3228-z

Yang, T. (2008). Factors affecting EFL teachers' use of multiple classroom assessment practices with young language learners. *English Teaching & Learning*, 32(4), 85-123.

Yin, M. (2010). Understanding Classroom Language Assessment through Teacher Thinking Research. *Language Assessment Quarterly,* 7, 175–94.

Yu-Ching, C. (2008). Elementary school EFL teachers' beliefs and practices of multiple assessments. *Reflections on English Language Teaching*, 7(1), 37-62.

Vitae:

Irini-Renika Papakammenou, PhD in Linguistics, Private Institute of Foreign Languages and Research Centre I. Papakammenou,, Greece. Her research interests cover various aspects of language testing and assessment (test washback, alternative methods of assessment, material development for assessment), language teaching and learning aspects, teacher development and online education.

Correspondence: renikapap@yahoo.com

OLGA KVASOVA, TAMARA KAVYTSKA &
VIKTORIYA OSIDAK (Ukraine)

ORCID: 0000-0002-1479-0811
ORCID: 0000-0002-1528-9439
ORCID:0000-0001-7304-3026

INVESTIGATION OF WRITING ASSESSMENT LITERACY OF UKRAINIAN UNIVERSITY TEACHERS

The rating process of students' writing has been a long-standing concern in L2 large-scale standardized and classroom-based assessment. Several studies have tried to identify how the raters make scoring decisions and assign scores to ensure validity of writing assessment. The current paper addresses the issue of writing assessment practices of Ukrainian university teachers, how they approach rating scales and criteria with an attempt to understand culturally specific challenges of teachers' writing assessment in Ukraine. To investigate the issue, this study employs the analysis of the survey results obtained from 104 university teachers of English. The survey consisted of 13 questions that provided insight into current practices in assessment of writing, such as frequency of assessment, use of rating scales, rater's profile, criteria of assessment, feedback and rewriting, training in assessment of writing.

The survey responses show that assessment in Ukraine is not regulated by common standard, and thus the approach to students' writing assessment is often intuitive. A frequent practice is that teachers tend to rely on errors – as observable features of the text – to justify their rating decisions, Consequently, by shifting focus onto the surface features of writing, grammar mistakes in particular, the teachers underrate such criteria as "register", "compliance with textual features" and "layout". Additionally, the data reveal contradictory findings about writing assessment literacy of the teachers questioned. Even though most teachers claim they apply scales while rating, many confess they cannot tell the difference between holistic and analytic scales. Moreover, the results indicate that feedback is not yet a meaningful interaction between a Ukrainian teacher and a learner. Therefore, the results of the study demonstrate the need for the improvement in writing assessment practices, which could be achieved through providing training and reorientation to help Ukrainian teachers develop common understanding and interpretation of task requirements and scale features.

Key words: *criteria for assessment, rating scales, teacher assessment practices, writing assessment literacy, writing tasks*

1. Introduction

Internationalization of professional and educational domains witnessed in the world in recent years has dramatically raised the significance of good writing skills in English as L2. As a result, most large-scale placement and proficiency tests include writing tasks as a hallmark of candidates' ability to express themselves in writing. Consequently, the issues of testing writing skills as well as candidates' preparation for taking standardized exams have been in the focus of international research in language assessment and pedagogy. Directed by standardized assessment objectives, exam preparation classes are conducted across the world to equip candidates with test taking strategies; specially trained teachers engage the prospective examinees in teaching-to-test.

Most learners, however, are still building their writing skills in L2 classroom settings. Here, both assessment and teaching are challenging demanding for teachers who had not been sufficiently trained in assessing writing and/or teaching to write. Moreover, in many contexts, teachers had not been taught to write in L2 properly themselves (Lee 2010, Crusan et al., 2016).

Teacher's ability to conduct fair measurement of learners' skills in their day-to-day work plays a crucial role in today's classroom. As Crusan et al. (2016) note: "Bad assessment practices can have a potent effect on students. The consequences of uninformed assessment can be losses for students in time, money, motivation, and confidence." (p.43.) Implementing writing assessments, the researchers maintain, teachers need to know how to score, grade, and make judgments about students' achievements; this is a laborious, weary, and time-consuming, as well as demanding activity.

Keeping in mind both national and international value of research into context-specific educational frameworks, we dedicate this paper to the investigation of teachers' literacy in assessment of writing in tertiary education in Ukraine. To this end, we explored the writing assessment literacy (WAL) by surveying testing practices of English teachers working in several universities across the country.

2. Literature review

Assessing writing has been mostly researched in the field of standard proficiency exams. Differ-
ent factors have been looked into such as the conceptualization of validity in assessing writing (Weigle, 2002), assessment criteria and their impact on scoring decisions made by trained raters (Weigle, 2002; Lumley, 2002), comparison of the existing rating scales and the empirical development of a scale based on discourse analytic measures (Knoch, 2007; 2009).

Much attention has also been paid to analysing the characteristics of raters from a number of different angles. Many scholars and researchers placed raters in the centre of writing performance assessment advocating that raters' educational background, language proficiency, rater experience and expertise in scoring, as well as personal traits were crucial to the objectivity of measuring writing skills (Johnson & Lim, 2009; Bijani, 2010; Lim 2011; Alaei et al., 2014). Lim (2011) provided the evidence of a special role of rater training and its impact on novice and experienced raters' scoring and its quality (p. 543). Examining the rater training procedures and the effectiveness of rater training programmes has been another area of interest to researchers (Attali, 2015; Bijani, 2010; Fahim & Bijani, 2011; Lim, 2011).

In spite of this bulk of research, rating has always been a subjective activity. This is primarily reflected in rater variance in scoring. Many studies highlight the determining role of raters' subjective perception of the evaluative criteria that can put validity of the test itself at risk. For example, Eckes (2013), through a synthesis study, provided a comprehensive review of the relationship between rater cognition and rater behavior in terms of severity/leniency. According to the results of the study "[a] rater considering a particular criterion as highly important, and other criteria as less important, would tend to score on this criterion more harshly than expected based on his or her overall severity and the overall criterion difficulty measure" (Eckes, 2013, p. 285). This suggests that personal experience of what a good writing is often interferes with interpretation of the scoring rubric. A similar bias is observed when raters score more leniently on the criteria which are thought less important by them. Also, raters may put different emphasis to various components of scale descriptors when they have to rate writings of particular level of overall proficiency.

Recently new perspectives have emerged in the field of the classroom-based assessment with regard to the ways of how teacher's writ-

ing assessment literacy and testing practices in classroom settings across the world are being conceptualised. Analysing the impact of classroom-based writing assessment on learners' achievements, raters' consistency (inter-rater reliability), raters' behaviour (severity and bias), and raters' interaction with different types of written texts (different types of essay, composition, etc.) have been but few areas of interest to many researchers (Cho, 2008; Lee, 2010; Ghanbari et al., 2012; Jeong, 2015; Crusan et al., 2016; Nemati et. al., 2017; Skar & Jølle, 2017; Mellati & Khademi, 2018; and Aslim-Yetis, 2019). This result is in agreement with other studies about the importance of teachers' understanding of rating scales and scoring criteria, which could be enhanced through proper training and teacher development (Hamp-Lyons, 2007, p. 1). This finding also adds evidence that the improved standards to classroom-based assessment is only possible as the result of vastly improved attention to teacher training in relation to assessment.

As a rule of thumb, apart from possessing overall literacy in writing assessment, teachers should share a common understanding of rating scales and be able to apply them in particular educational contexts. In this regard, Ghanbari et al. (2012) has pointed out the importance of context-based rating scales, and teachers' ability to tailore them wherever possible to their teaching/assessing needs.. He contends that "… [a] local rating scale takes into account the particularities of each assessment context [which] would lead to more valid outcomes. Such a proposal … is justified and supported by both validity argument of Weir (2005) and critical discussions in the field." (p.97).

Given the primary importance of teachers' writing assessment practices, their understanding of rating scales and criteria, as well as personal beliefs about scoring and feedback effects on learners, the current study aims to investigate the context-specific context of Ukrainian higher education. The research questions of the study are:

– What and how is students' writing assessed today?

– What are the possible gaps in WAL of Ukrainian university teachers? How can these gaps be bridged?

3. Research methodology
1) A survey of WAL of Ukrainian university teachers

The questionnaire used in the survey included 13 questions aimed to elicit current practices and training in assessment of writing (received and needed). Some questions were selected response, others were ordering. The majority of the questions had an open-ended option "other" thus encouraging teachers to share their experience in a more detailed way.

a) Frequency of assessment of writing and text types most/least frequently scored. Q 1 was aimed to elicit how frequently teachers generally rated their students' writing. By asking Q2 we intended to find out what writing tasks teachers rated most and least frequently.

b) Use of rating scales. Q 3 was concerned with the authorship of the writing tasks most and least frequently administered; Q 4 asked if teachers used any rating scales while scoring; Q5-Q 6 were aimed to specifically find out what kind of rating scales (analytic/ holistic) the respondents use and who developed most and least used rating scales.

c) Rater's profile. In response to Q7 the teachers were to provide information on their performing rating, and how harsh/lenient they were self-evaluated as raters.

d) Criteria of assessment. Q8 was another ordering task aimed to evaluate *most and least often evaluated* aspects of writing. Q9 aimed to find out if teachers explained the criteria of assessment to students and, if yes, whether this preceded or followed the writing task.

e) Feedback. Rewriting. Q 10 -11 were to collect the responses about the format of feedback the teachers most and least often provided to learners, and if they typically required rewriting of papers.

f) Training in assessment of writing. Q 12 was to ascertain what training in assessing writing the respondents received, and Q13 was to order forms of training from *"most needed"* to *"least needed"*.

2) Participants in the study

The questionnaire was responded to by 104 university teachers of English, 68 of whom taught ESP, whereas 36 were teachers of GE and EAP working for linguistic and teacher-training universities. All participants were female; they represented nine regions of Ukraine. Participation in the survey was voluntary. The survey was conducted in February – March 2018.

4. Findings and discussion

Below we provide the data obtained via the survey as well as analysis and interpretation of the responses.

a) The answers to the first group of questions indicated that university teachers were consistently engaged in assessment of writing. In the majority of instances (53%), the respondents performed rating procedures once a month; 22% assessed writing once in two months, 10% twice a month, and 5% once a term. The variance in frequency depended on the curriculum the respondents taught. Obviously, in schools of linguistics writing tasks were set most frequently – twice or once a month. With regard to the most frequent types of writing, the three top-ranked ones were identical in the three settings, namely essays, letters, and reports, followed by proposals in ESP, articles and stories, and abstracts, summaries and reviews in GE and EAP respectively. Given that even the mostly evaluated genres exist in a variety of subgenres, teachers should be knowledgeable in generic as well as genre specific conventions and rank them as most or/and least significant in order to assign fair scores.

b) In respect of the authorship of the administered writing tasks and rating scales used, we revealed the following. In less than a half of incidences (48%) the teachers used the writing test tasks offered by the authors of the coursebooks, which might be explained either by the lack of such tasks or their irrelevance to the teaching/ assessment situations. Presumedly, this is the reason why 53% of respondents prepared tasks mostly by themselves and 34% collaborated with other faculty members.

The data revealing the use of rating scales were somewhat contradictory. When answering if they used scales while rating, 79% of respondents claimed they did, with only 4% conceding their ignorance in that matter. When further asked to specify what kind of scale the teachers used – holistic or analytic, it appeared that not 4% but 28% of respondents were not aware of the concepts whatsoever. As for the rest of the respondents, their preferences for kinds of scales were almost equal: 32% used analytic and 33% used holistic scales.

Interestingly, despite the revealed misconception of rating scales, 42% of respondents stated that they most often developed the scales individually, 25% collaborated in scale design with other teachers and 17% used the scales developed by other teachers. Only 39% and 28% of respondents most frequently resorted to the use of scales offered by the coursebooks or examination systems, such as Cambridge ESOL or IELTS, respectively. The reason for teachers' disregarding the scales developed by courses book authors might lie in the absence of such in the provided teacher's books; the scales developed by testing experts might look arcane to teachers, which might lead them to relying on own experience and /or intuition.

c) Since the contradictory data about the use of rating scales questioned the validity of scoring based on teacher-designed rating scales, it is interesting to look at teachers' self-evaluative responses. The majority of respondents (56%) regarded themselves fair raters whereas only 4% admitted they were harsh, the rest 40% thought they were lenient but wished they had been harsher raters.

We also managed to identify some scoring practices. According to the majority of the respondents (39%), they basically gave two grades – "one for content and one for accuracy"; 37% claimed that they wrote "a short commentary on each work to make the rating clearer to students"; 24% tended to give one grade reflecting the general impression of the written text because for them, "content is more important than accuracy (as long as the message is conveyed)". The above responses seem to reflect teachers' commitment to provide students with detailed feedback on their writing; however, the teaching experience of this paper's authors cast doubt on the feasibility of writing comments on each student's paper under the workload carried out by teachers.

Similarly, based on our teaching experience and being aware of the 'penalizing-for-any-error' approach adopted by many Ukrainian teachers, we aimed to find out what kind of errors – grammar, vocabulary or spelling – played an essential role in teachers' rating of papers. Thus, 43% of the respondents revealed that they most frequently deducted points for wrong choice of words, 37% were harsh at grammatical errors, 20% claimed they penalized writers for spelling mistakes.

d) If respondents were mostly concerned with deducting points for language errors, then what criteria did they find worth using while rating? In the survey, the teachers were offered a list of

criteria which they had to order as the most and least frequently used in their practice. The graph below shows the ranking of the criteria.
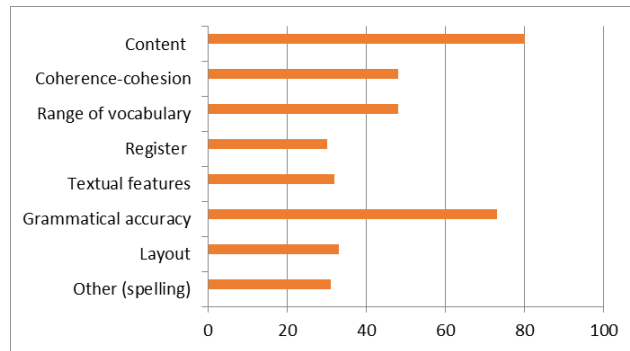

Table 1. Criteria for assessment of writing

As is seen in the graph, "Grammatical accuracy" was the only criterion apart from "content" that the teachers believed was worth considering. Equally important, though ranked considerably lower, were criteria "Coherence and cohesion" and "Range of vocabulary". As for the latter index, it contradicts the claim that the respondents were more tolerant to vocabulary than to grammatical inaccuracy. Criteria "Register", "Compliance with textual features", "Layout" and "Spelling" were almost equally underrated by the respondents. These data once again indicate that teachers were more inclined to evaluating the most vivid to them and easier assessable aspect of writing, such as grammar.

The significance of criteria "Content" is deep-rooted in the national tradition of teaching writing both in L1 and L2. Yet, the use of this criterium might become an issue for concern as it introduces a potential bias of background information (Weigle, 2002, p.95). As the result, a writing task will disadvantage those students who may have no advanced knowledge on a topic of a writing task (Chan, 2013; Weigle, 2002, p.52). The underrating of "Register", "Compliance with textual features", and "Layout", in our view, could have been smaller, in case more teachers had been familiar with and used the rating scales offered by international testing systems. There, unlike in teacher-applied criteria, they occupy a notable place.

The awareness of criteria is generally viewed as necessary for raters and desirable for test takers. According to the respondents, the overwhelming majority of them (89%) tended to inform students about the assessment criteria before their writing, which testified to the test transparency.

e) No less important was to examine the practices of feedback provision to students. Commendable were the practices that received top indeces: 54% of providing individual (written) and 50% of providing individual (oral) feedback to students.

As is known, individual feedback tends to perform corrective function more efficiently (Brookhart, 2011; Nemati et al., 2017). Making general comments on the work of the class in general appeared the next preferable form of feedback (36%); nearly the same percentage of respondents (34%) indicated that they provided detailed feedback on each work before the whole class, which is less desirable. The negative practice of providing feedback on students' written work on their request only was reported by 30% of respondents. None at all or late feedback is viewed as unacceptable for "feedback is of no value if it arrives in the students' hands too late to do anything with it", as Coombe et al. maintain (p.xxvi).

The role of feedback is all the more valuable if it prompts rewriting a failed text which is considered an effective remedial activity. Students' willingness to rewrite ascertains their acceptance of teachers' corrective feedback, it also reflects students' motivation to improve their writing. Ideal then is the practice when teachers provide the opportunity of multidrafting a text.

Does such practice exist in Ukrainian higher education? The data of the survey do not provide a definitive answer: 15% of teachers claimed they always set rewriting to their students whereas 18% never practiced that. In 50% of incidences teachers asked students to improve their written texts in some cases only; 14% of teachers welcomed rewriting only by the students wishing to do that. On the whole, the greater part of the respondents who practiced occasional, regular or volunteered rewriting, suggests that they appreciated that activity as beneficial to learners and would probably resort to it more frequently if circumstances had permitted.

f) The final part of the survey meant to elicit teachers' perceived needs in training in writing assessment.

As the bars indicate, in the overwhelming majority of cases (73%) Ukrainian university teachers were trained in assessment of writing via self-study; quite similar are figures indicating that they were taught via own experience (55%) or through participation in workshops (48%). Webinars as a source of WAL were mentioned only by 21%. The
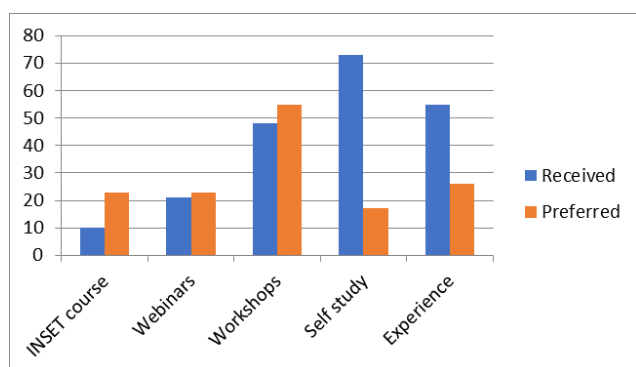
Table 2. Received and preferred training in assessment of writing

smallest percentage of respondents (10%) said that they owed their WAL to formal education – INSET courses. Quite opposite are the indices of preferable training. Interestingly, the respondents would not wish to further rely on their self-study (17% vs 73%) and experience (26% vs 55%) but want to be guided in their training by experts within INSET courses (23%) and webinars (23%) while the most preferred method of building WAL was workshops (55%).

The evidence collected via the survey provides interesting insights about Ukrainian university teachers' assessment practices as reflecting their WAL. According to the data, respondents were engaged in administering and assessing students' writings on a continual basis, however, only an insignificant percentage of the respondents claimed they had received a formal training in WAL building. As a result, while rating students' papers, teachers have to rely mostly on their competence in teaching how to write the most relevant for their settings texts. This competence is generally acquired by them during the pre-service training and further fostered in teaching experience; similarly, teaching experience and self-study account for WAL building in the most typical practices surveyed.

Assessment of writing, as responses suggest, is not regulated by any common standards which are non-existent either in the national or local dimensions. The selection of criteria of assessment is therefore at the teachers' discretion; the respondents in the survey most frequently evaluate content and grammatical accuracy of students' writings. Such practice should be abandoned overall, according to Truscott (1999) and Keh (1993). Truscott (1990) also put the view, citing numerous studies, that grammar correction does not bring the desirable improvement to writing accuracy and literacy; and grammar mistakes usually disappear with the improvement in stu-

dents' writing proficiency. The research of Chan (2013) suggests that content cannot be judged harsh since it reflects subjective vision of the world. As for the severity of assessing grammatical accuracy, Kim & Gennaro (2012) argued that non-native speaking raters have always been harsher raters than the native speakers.

Introduction and use of such criteria as coherence and/or organisation are fraught with difficulties even in the case of training raters for standardized assessment. Knoch (2007) suggests that these difficulties might result from "vague descriptions of coherence" possibly due to "rather vague nature of coherence itself" (p. 109). Selection / use of assessment criteria and methods of scoring could optimize teachers' if rating scales had been introduced in the assessment of writing practices.

Today, the predominant percentage of the respondents suggest, they rely on rating scales in their rating, however, the data cannot be considered reliable due to the revealed misconceptions of holistic and analytic rating scales. This inconsistency calls for substantial rater teacher training either in the use of ready-made scales offered by international exam providers, in tailoring them to local curricula or developing local scales. In the last case, the significance of teachers' linguistic competence, as well as their competence in teaching/assessment pedagogy is of primary importance.

On the whole, the study conducted implies that it is necessary to provide university teachers with training in understanding theoretical underpinnings of assessing writing and applying the theory in practice thus implementing good testing practices and ensuring their positive washback on learning. The decision to organize such training rests on educational authorities; the development of ways to enhance WAL of university teachers shapes the perspectives of future research.

5. Conclusion
A considerable bulk of research is devoted to the investigation of teachers' writing assessment practices to understand how grades can become a better reflection of students' learning. This study was an attempt to broaden our understanding about WAL of university teachers in Ukraine by using a survey data. Collected results showed that university teachers in Ukraine are actively and regularly engaged in writing assessment. Also, most respondents preferred preparing writing tasks by themselves or with other colleagues to using tasks

offered in a coursebook. Besides, almost half of the teachers, who participated in the study, developed rating scales themselves. This might be an indication that teachers believed they are qualified enough to develop a test / rating scale and / or that they view test tasks / rating scales offered in coursebooks / by testing experts as irrelevant for their classroom assessment purposes.

Yet, a staggering percentage of teachers recognise the need for enhancing their general assessment literacy as they believe that their mostly self-acquired competence in WAL is not satisfactory for conducting assessment. The study collected enough evidence of teachers' inadequate training such as misconceptions of rating scales; penalising for any error approach; overviewing timely and effective feedback in enhancing students' performance etc. Therefore, the next step that should be taken is help university teachers in Ukraine enrich their WAL through training guided by experts.

References

Alaei, M.M., Ahmadi M., & Zadeh N.S. (2014). The Impact of Rater's Personality Traits on Holistic and Analytic Scores: Does Genre Make any Difference too? *Procedia – Social and Behavioral Sciences 98*, 1240 – 1248. Available online at www.sciencedirect.com

Aslim-Yetis, V. Evaluating Essay Assessment: Teacher-Developed Criteria versus Rubrics. Intra/ Inter Reliability and Teachers' Opinions. Croatian Journal of Education. Vol.21; No.1/2019, pages: 103-155. https://doi.org/10.15516/cje.v21i1.2922

Attali, Y. (2016). A comparison of newly-trained and experienced raters on a standardized writing assessment. *Language Testing,* Vol. 33(1) 99–115. DOI: 10.1177/0265532215582283.

Bijani, H. (2010) Raters' Perception and Expertise in Evaluating Second Language Compositions. The Journal of Applied Linguistics Vol. 3, Issue 2, p.69-89. Retrieved from Researchgate.

Brookhart, S. M. (2011). Educational assessment knowledge and skills for teachers. *Educational Measurement: Issues and Practice, 30*(1), 3–12. doi: 10.1111/j.1745-3992.2010.00195.x.

Chan, S.H.C. (2013). Establishing the validity of reading-into-writing test tasks for the UK academic context / A thesis submitted to the university of Bedfordshire, 376p.

Cho, D. (2008) Investigating EFL Writing Assessment in a Classroom Setting: Features of Composition and Rater Behaviors. THE JOURNAL OF ASIA TEFL Vol. 5, No. 4, pp. 49-84.

Coombe, C., Folse, K., Hubley, N. (2007). *A practical guide to assessing English language learners.* Ann Arbor, Michigan: University of Michigan Press.

Crusan, D., Plakans, L., & Gebril, A. Writing assessment literacy: Surveying second language teachers' knowledge, beliefs, and practices. Assessing Writing – Volume 28, April 2016, Pages 43–56. https://doi.org/10.1016/j.asw.2016.03.001

Eckes, T. (2012). Operational Rater Types in Writing Assessment: Linking Rater Cognition to Rater Behavior. *Language Assessment Quarterly*, 9: 270–292. DOI: 10.1080/15434303.2011.649381

Fahim, M., & Bijani, H. (2011). The effects of rater training on raters' severity and bias in second language writing assessment. *Iranian Journal of Language Testing*, *1*(1), 1-16.

Ghanbari, B., Barati, H. and Moinzahed, A. (2012). Rating Scales Revisited: EFL Writing Assessment Context of Iran under Scrutiny. *Language Testing in Asia*. Volume two, Issue one. 2012, p. 83- 100.

Hamp-Lyons, L. (2007a). The impact of testing practices on teaching. In *International handbook of English language teaching*, 487-504. Springer, Boston, MA

Hamp-Lyons, L. (2007b). Worrying about rating. *Assessing Writing*, *12*(1), 1-9. https://doi. org/10.1016/j.asw.2007.05.002.

Jeong, H. (2015). Rubrics in the classroom: do teachers really follow them? Language Testing in Asia 5:6. DOI 10.1186/s40468-015-0013-5.

Johnson, J. S., & Lim, G. S. (2009). The influence of rater language background on writing performance assessment. *Language Testing*, *26*(4), 485-505. http://dx.doi. org/10.1177/0265532209340186

Keh, C. L. (1990). Feedback in the writing process: a model and methods for implementation. *ELT Journal,* (44)4, 294-304.

Kim, Ah-Young and Gennaro, di Kristen. (2012). Scoring Behavior of Native vs. Non-native Speaker Raters of Writing Exams. Language Research 48.2, 319-342.

Knoch, U. (2007). 'Little coherence, considerable strain for reader': A comparison between two rating scales for the assessment of coherence. Assessing Writing 12 (2007) 108–128. Available online at www.sciencedirect.com.

Knoch, U. (2011). Rating scales for diagnostic assessment of writing: What should they look like and where should the criteria come from? *Assessing Writing, 16,*81-96.