JEELS
(Journal of English Education and Linguistics Studies)
P-ISSN: 2407-2575 E-ISSN: 2503-2194
https://jurnalfaktarbiyah.iainkediri.ac.id/index.php/jeels

KURDISH EFL TEACHERS' PERSPECTIVES ON ASSESSMENT AT UNIVERSITY LEVEL: FACTORS, CHALLENGES, AND WASHBACK

***Ivan Hasan Murad[1]; Sanan Shero Malo[2]**
[1,2]English Language Department, College of Humanities, University of Zakho, Kurdistan Region-Iraq
*ivan.murad@uoz.edu.krd sanan.malo@uoz.edu.krd
(*) Corresponding Author

Abstract: Assessment constitutes a fundamental component of educational process, shaping teachers' instructional practices. However, its efficacy largely depends on the way teachers perceive assessment. Thus, this research article investigates Kurdish university-level English as a Foreign Language (EFL) teachers' perspective on assessment in terms of the factors that influence their choice of assessment methods, their challenges in assessment, and the washback effects of assessment on students' learning and study habits. Employing a quantitative approach, the study utilized a Likert-scale survey to collect data from a random sample of 75 teachers from 11 public universities in the Kurdistan Region of Iraq. Data were analyzed using

[1]**Citation in APA style:**
Murad, I. H., & Malo, S. S. (2025). Kurdish EFL teachers' perspectives on assessment at university level: Factors, challenges, and washback. *JEELS (Journal of English Education and Linguistics Studies), 12*(2), 1029-1057.
DOI: 10.30762/jeels.v12i2.6562

Submission: July 2025, Revision: October 2025, Publication: November 2025

the Statistical Package for Social Sciences (SPSS) version 25. The findings revealed several factors influencing teachers' assessment choices, including strict institutional regulations, large class sizes, and students' growing reliance on AI tools. The study also found that teachers face numerous challenges, such as detecting AI-generated work, limited access to technological resources, and balancing formative and summative assessments. Moreover, the results indicate that assessment practices generate both positive and negative washback, influencing students' motivation, effort, and learning approaches. The study highlights the importance of aligning assessment practices with Kurdish EFL higher education principles to promote effective learning outcomes.

Keywords: *Assessment, challenges, factors, washback, Kurdish EFL teachers, University-level setting*

INTRODUCTION

Assessment plays a crucial role in English as a Foreign Language (EFL) education, shaping both teaching practices and learning outcomes. At the university level, particularly in EFL context where students look forward to achieving high level of linguistic and academic competence, assessments are designed to measure their language proficiency, critical thinking, and communicative competence (Bouckaert, 2023). However, the effectiveness of these assessments is contingent upon the way they are perceived and implemented by instructors. Therefore, EFL teachers possess a range of beliefs and perspectives on assessment that are crucial which influence not only the design and execution of assessment, but also teachers' broader educational impact and the quality of students' learning, including the washback effect (Yin, 2010). Understanding these perspectives can help identify gaps between assessment policies and classroom realities, ultimately leading to more effective evaluation practices.

Throughout history, the focus of assessment has witnessed changes. In 1990, assessment focused on the measurement of what students have learned. Back then, according to Brown (1990), assessment was defined as a series of evaluations and measures utilized to check the achievement of learners. It involved collecting and analyzing data on the students' achievements and learning. It also measured the improvement of students, the teaching methods, and the students' motivation. Following this, Allen, (2004) defined it as the systematic process by which a documentation of students' data regarding their knowledge, attitudes, aptitudes, skills, and beliefs for the purpose of refining educational programs and improving students' learning. According to the Great Schools Partnership (2015), the term assessment refers to the wide variety of methods or tools that educators use to evaluate, measure, and document the academic readiness, learning progress, skill acquisition, or educational needs of students. The role of assessment continues to progress with the evolvement of the technological tools utilized in the educational assessment. Assessment, therefore, is defined as a dynamic, learner-informed procedure that incorporates digital tools, student voice, and real-time analytics to support equitable, personalized, and lifelong learning pathways (Darling-Hammond et al., 2022).

In the Kurdistan university-level EFL context, assessment has witnessed many changes and development towards the better based on the systems of education which have been utilized, the Bologna Process as an example, to catch up with the modern and globalized world of knowledge. This process is defined by Wächter (2004) as "a unified initiative involving governmental bodies, academic institutions, educators, students, stakeholder organizations, employers, and quality assurance agencies aimed at systematically and cohesively restructuring higher education frameworks across Europe" (p. 22). According to Ade (2021), the process is an intergovernmental collaboration between 48 countries in Europe with a focus on improving the quality of HE in terms of facilitating the mobility and coordination between these countries which involves public

authorities, universities, students, educators, stakeholders, international entities, quality assurance bodies, and other institutions. In terms of assessment process, unlike the traditional annual system of education in which there was a heavy dependence on tests, this system encourages a range of assessment strategies, methods, and workload that encourage learner-centered approaches such as presentations (in each semester), reports, assignments, quizzes, projects, homework, mid-term exams, final exams etc.

Assessment is an umbrella concept that encompasses both testing and a wide array of evaluation strategies. Scholars argue that the concept of assessment extends beyond the traditional forms of assessment, including testing, as it involves more than testing for the purpose of providing a holistic evaluation of the students' outcomes and performances (Brookhart & Nitko, 2019; Chappuis & Stiggins, 2017; Wiliam, 2018), especially in higher education. According to several studies, university-level assessment should be a systematic and a continuous process in nature, with clearly defined objectives, assessment methods, and prompt and consistent feedback (Chappuis & Stiggins, 2017). In addition, as it is argued by William (2018), assessment must be a complementary part of teaching with formative and summative strategies aiding in supporting quality learning and academic progress.

The implementation of the assessment methods by teachers is influenced by a number of factors. Building on Kozma's (2003) framework, Fulmer et al. (2015) came up with a three-tiered model for categorizing the influences of assessment practices, which are used by several researchers (Deneen et al., 2019; Ma, 2023). The current study also addresses the same classification system as factors influencing teachers' choice of assessment methods, examining factors at micro, meso, and macro levels.

At the micro level, personal and classroom factors have their own influence. These include teachers' assessment literacy, self-efficacy, beliefs, and experience; and contextual elements like class size, technology access, and classroom dynamics (Black & Wiliam, 2018; Yan

et al., 2022). The meso level encompasses institutional influences such as institutional policies, administrative guidelines, and stakeholder expectations (Ma & Bui, 2021). Macro-level elements involve broader policy scopes, particularly national and international assessment norms, high-stakes testing regimes, and their washback effects (Ma & Bui, 2021; Wilson, 2024). These interconnected levels collectively influence teachers' assessment choice and implementation strategies.

According to Cheng and Curtis (2010), due to administrative and accreditation requirements and pressures, institutions and educational bodies ought to abide by prescribed assessment guidelines. Educational institutions establish departmental guidelines and institutional standards to ensure that assessment practices are consistent, valid, and aligned with curricular goals. These frameworks provide teachers with clear criteria for selecting appropriate evaluation methods, thereby promoting fairness and comparability across courses and programs. However, teachers might think that one size does not fit all, and feel that they need to choose different assessment methods from those assigned by the department or institution. In addition, in educational contexts, particularly in EFL settings, standardized testing is widely spread and more favored due to administrative requirements and accountability pressures (Cheng et al., 2007). This could limit teachers' freedom in implementing certain methods of assessment such as formative or performance-based assessments that may better address diverse student needs. Consequently, teachers may feel obliged to follow prescribed assessment policies, which sometimes conflict with their professional judgment and preferred pedagogical approaches.

Assessment methods that require less time to be administered are often more favored by teachers especially those who have classes with large number of students (Hamp-Lyons, 2017). Such teachers might tend to favor particular methods of assessment such as multiple-choice questions, short answer questions, or even automated scoring tools or those methods that represent the true level of their students such as essay-writing, oral presentations, and project assignments

(Fulcher, 2020). This factor also impacts the quality of the assessment in terms of the validity of the assessment method, as overemphasis on reliability may compromise validity (Erlinawati & Muslimah, 2021).

The availability of resources alongside teachers' in-service training has a big role in the provision of quality assessment practices that reflect the real performance of the students. According to Giraldo, (2021), the availability of limited resources leads teachers to avoid utilizing modern assessment methods such as computer-based assessments or performance-based ones. Moreover, the insufficient teaching and working experience, especially in terms of designing assessments or implementing and administering them, leads to focusing on the ones that are familiar but could be outdated (Malone, 2022).

The scoring criteria of particular assessment methods also influence the choice of the method being utilized. This is because the time-consuming factor in terms of scoring might influence EFL teachers' willingness to choose particular assessment methods (Brookhart, 2017). Moreover, student's reliance on AI in obtaining read-made answers has put much pressure on teachers to reconsider the written assessments. This has driven them to focus more on assessments that require in-class writing, oral presentations, or preparing projects that require more complicated tasks (Warschauer, 2020).

EFL teachers face challenges in the assessment process, ranging from technological disruptions to institutional constraints. The rise of AI tools like ChatGPT and Grammarly has made authentic assessment more complicated. This is because teachers struggle to evaluate genuine language and knowledge proficiency when assignments may be AI-generated (Warschauer, 2020). To address this, as it is already mentioned above, educators increasingly adopt in-class writing, oral presentations, or project-based assessments, though these methods demand additional time and effort (Perkins et al., 2023). Large class sizes further exacerbate assessment difficulties, forcing teachers to rely on less nuanced methods like multiple-choice tests rather than essays

or speaking tasks, which limits formative feedback and hinders learning outcomes (Hamp-Lyons, 2017; Tsagari & Vogt, 2017). Students' level and performance that might be inconsistent due to varying prior knowledge and motivation complicates assessment, requiring differentiated strategies that standardized tests fail to achieve (Brown & Abeywickrama, 2019; Kunnan, 2018). In addition, limited access to technological resources in many EFL classrooms, especially in Kurdish EFL classrooms, restricts the use of digital assessment tools, compelling teachers to rely on traditional paper-based methods that may not fully measure communicative competence (Giraldo, 2021; Shadiev & Hwang, 2020).

Institutional guidelines constrain EFL teachers by prioritizing summative over formative assessment methods, stifling innovation in methods of assessment such as authentic ones like portfolios or peer evaluations (Cheng et al., 2007; Fulcher, 2020). Making balance between formative and summative assessments continues to be a persistent challenge for EFL teachers, as challenges such as time constraints and curriculum demands often lead EFL teachers to opt for high-stakes testing, reducing opportunities for meaningful feedback and authentic assessment (Black & Wiliam, 2018; Popham, 2018).

Grading in some particular modules such as writing and speaking skills, initiates a sort of bias which causes problems for teachers despite the use of assessment strategies that mitigate bias, highlighting the need for better rater training and assessment literacy (Weigle, 2020; Malone, 2022). Lastly, the administration workload of preparing, administering, and grading assessments causes teacher burnout. This is because much paperwork detracts from student engagement and provision of quality assessment and teaching (Bachman & Damböck, 2018; Bailey & Brown, 2022). These challenges underscore the need for support, professional training and development, and flexible assessment strategies to enhance EFL teaching efficacy.

The washback effect of assessment in EFL contexts profoundly shapes students' learning behaviors, language development, and study

habits for it is considered as a reflection of the way assessment impacts them (Hattie & Timperley, 2007). Well-designed assessments have been confirmed and evidenced to promote language proficiency by offering students continuous academic support for the purpose of skill reinforcement and knowledge development. Frequent, low-stakes assessments are particularly impactful in promoting memorization and fluency, provided that they are detailed, constructive feedback fosters metacognitive awareness and self-regulated learning (Black & Wiliam, 2018; Hattie & Timperley, 2007). The diagnostic function of assessment enables both students and teachers to identify specific strengths and weaknesses, allowing for targeted interventions that address individual language needs (Alderson et al., 2015; Kunnan, 2018). Moreover, other forms of authentic assessments such as project-based tasks, critical essays, and many others encourage higher-order thinking and scrutinizing skills, analytical reasoning, and cognitive engagement that make students confident in their skills and knowledge, moving beyond rote memorization to cultivate deeper understanding and concept development (Brookhart, 2017; Shadiev & Hwang, 2020).

According to the best knowledge of the researchers, not many studies have been conducted to address the aims and themes of the current study especially in the context of Kurdish EFL education at university level leaving a gap in literature that needs to be addressed.

Regarding the factors that influence university-level teachers' choice of assessment methods, a recent study was conducted by Williams-McBean (2025) to investigate factors affecting teachers' choice and use of assessment. The study employed a mixed-method design employing a survey administered to 1088 Jamaican school teachers and interviews with 32 English teachers. The study revealed that factors such as beliefs and knowledge, institutional policies, and international assessment norms, high-stakes testing regimes, and their washback effects have a huge role in teachers' assessment choice.

Latif and Wasim (2022) conducted a study to understand teachers' assessment literacy by investigating their personal beliefs, conceptions, and theories regarding the assessment methods. The

study was qualitative in nature and twelve teachers from three different institutions took part in it. A thematic approach was followed to analyze the data gathered from the semi-structured interviews. Results revealed complexity, diversity, and uncertainty in teacher beliefs. The study also found out that contextual and institutional policies and regulations have an undue influence on teachers' choice of assessment methods.

In their study, Tanberkan, et al. (2024) investigated the impact of AI on assessment and evaluation approaches in education. The objective the authors aimed at was to check both the benefits and ethical challenges posed by AI technologies for students and teachers. Using a literature review as the primary data collection method, the study highlights opportunities such as individualized learning and efficient assessment, but also stresses risks, especially for students lacking AI ethical literacy who may resort to unethical practices.

In another study, Hussain et al. (2019) explored the classroom assessment practices of Grade 10 English teachers in Khyber Pakhtunkhwa, Pakistan, focusing on both the challenges and opportunities faced at the secondary level. The study employed a quantitative approach and data were collected via a questionnaire from 235 teachers in Districts Mardan and Swat. The study revealed key challenges including time constraints, having big class sizes in terms of number of students, lack of assessment training, and concerns over the psychometric soundness of their tools.

Albondoq (2023) did a study to investigate the washback effects of assessment on English language learning and teaching. This study stressed the crucial role of feedback for learners and their academic progress and explored washback's impact on students' language learning skills, motivation, and teachers' instructional techniques. A questionnaire was employed to gather data from 42 English language teachers and 200 students across the English and Translation departments. Results demonstrated that washback positively impacts students' language learning skills, development, and motivation.

However, individualized feedback was seen not to have been possible due to heavy academic loads.

A study by Omar (2020) investigated the effect of examination on teaching and learning. The study was conducted in Kurdistan and involved high school teachers teaching English language. The researcher employed a survey of two parts, one collecting demographic information and the other part consisted of 30 items on washback effect of examinations. Results from the study revealed that the examination had a negative washback on both teaching and learning. Teachers tended to teach to the test and students' focus shifted from learning the material to studying for the test only.

In spite of the fact that assessment has a critical role in EFL education, there is a gap in understanding how university-level Kurdish EFL teachers perceive and navigate the assessment in terms of factors that influence their choice of assessment methods, challenges they face in assessment, and the washback effect of assessment on students' learning and study habits. Existing studies have explored washback effects in standardized testing (Cheng, 2005) or school-based assessments (Qi, 2005 as cited in Arthur & Partey, 2023), but fewer have examined university settings where assessment practices may differ due to academic autonomy and diverse student needs. Additionally, while challenges in assessment implementation have been documented (e.g., time constraints, lack of training), little research has systematically investigated the institutional, cultural, and pedagogical factors that influence the choice of assessment methods from teachers' perspectives. By addressing these issues, the research aims to contribute to more informed assessment policies and teacher support mechanisms.

Therefore, the current study holds significant implications for university-level Kurdish EFL teachers and also other foreign EFL teachers for it contributes to the growing body of research on key aspects in assessment and EFL education. Through examining the factors that influence teachers' choice of assessment methods, their potential washback effects, and the challenges they encounter in the

assessment process, this research tries to develop a more nuanced understanding of the way assessment can be enhanced for the purpose of ensuring deeper learning.

METHOD

Research Design

The current study employs a quantitative approach in examining the university-level Kurdish EFL teachers' perspectives of assessment. It particularly studies their perspectives of the factors that influence their assessment choice, the challenges they face in assessment, and the potential washback effect of assessments on EFL student learning and study habits. The current study employed a quantitative research instrument in the data collection process. The process began with the distribution of a survey (see the Appendix) especially designed for the purpose of the study. According to Mertens (2019), utilizing quantitative approach helps in capturing the breadth of the topic under investigation.

Participants

The participants involved in this study consisted of 75 Kurdish EFL university-level teachers representing 11 public universities in the Kurdistan Region-Iraq (KRI). These universities include: University of Zakho, University of Duhok, University of SalahAldin, University of Sulaimani, University of Halabja, University of Soran, University of Garmian, University of Charmo, University of Koya, University of Raparin, and University of Acre for applied sciences. Such a diverse representation enriches the quality of the data being collected. Bryman (2016) states that for uncovering heterogeneous and effective experiences, collecting data from a wide range of samples is crucial. Moreover, it ensures multiple perspectives, improves the credibility of the data, and makes sure the collected information is representative (Creswell & Creswell, 2018). A random sampling technique was used in this study for it ensures representativeness of the subjects of the

study (Cohen et al., 2017), and gives all the population an equal chance to participate in the study (Fraenkel et al., 2012).

Data Collection Instrument: preparation and data collection process

A structured survey of a quantitative components utilizing Likert Scale (1=strongly disagree to 5=strongly agree) specially constructed as a part of a PhD thesis main instrument. As a total, the survey included 24 detailed items. Each 8 items of the survey represented one category: factors influencing EFL university-level teachers' choice of assessment methods, EFL teachers' challenges of assessment, and the washback impact of assessment. After the survey was constructed, it was given to a panel of jury consisting of 10 members specialized in English applied linguistics, for the purpose of evaluating and validating the items of the survey in terms of validity (face validity, content validity, and construct validity). According to Lieberman et al. (2016) and Creswell and Creswell (2018), to enhance the quality of any research instrument that has been constructed in educational research, it should be presented to a panel of jury of experts for the purpose of assessing whether or not the items adequately cover the intended content domain, ensuring its validity and reliability.

Based on the evaluation and feedback from the jury members, the majority of the items were edited as they were long and vague in terms of language. Then, the link for the survey was sent to 15 Kurdish EFL university-level teachers for piloting purposes. It is argued that piloting a data collection instrument increases its validity and reliability and boosts its practicality (Chhetri & Khanal, 2024). Moreover, Cohen et al. (2017) state that any instrument used in social science must be subject to piloting for the tool to be valid and reliable. After 15 days of the first piloting phase, a second round of piloting was conducted to eliminate all the confusions and misunderstandings of the items included in the survey. After this, the survey was ready for the data collection process. Based on the close connections that the staff of universities in KRI have, the survey was sent to colleagues from 11 public universities as Google Form link, which was shared with their

staff members for completion. These colleagues were either Head of English Language Departments or Directors of Language Centers. In less than two weeks, a total number of 83 responses were received. After critically reviewing the received responses, it turned out that 8 responses were not practical for 3 of them had straight neutral responses and 5 of them had missing data. Therefore, a total number of 75 valid responses were dealt with in this study.

Data Analysis

The data collected for this research study were inserted and analyzed employing the Statistical Package of Social Science (SPSS) version 25. The main purpose of this analysis was to measure participants' responses to the items of the survey by exploring the descriptive and inferential statistics. In order to determine the mean value and standard deviation for each survey item, the descriptive statistics were employed which provided an overview of central tendencies and the responses variability degree. Then, inferential analysis was conducted using one-sample t-test. This test was used to find out whether the observed mean values for the items of the survey significantly differed from the hypothetical value. In order to assess the level of the statistical significance, the t-test and corresponding p-values were computed. For determining the statistical differences, $p < 0.05$ was used as a threshold.

Ethical Considerations

To ensure confidentiality, the researchers have clearly indicated the intentions of the study including the aims and objectives. It has also been made clear that participation is voluntary and that all responses will strictly be kept confidential. It is important to make clear to the participants the true intentions of the study and that their responses will be kept strictly confidential (Kang & Hwang, 2023). Once the participants realize that their information and data are kept confidential, they tend to become more supportive in terms of data

they provide. Sherif (2018) confirms this and argues that when sound data collection process in terms of ethicality is used, rich and valid data can be gained.

FINDINGS

After the second piloting phase of the survey, the reliability of the items of the survey was checked. Cronbach Alpha was employed to check the reliability of the items of the survey. The utilized survey in this study consisted of three categories and each category consisted of eight items. Table 1 shows the results of the reliability check utilizing Cronbach's alpha.

Table 1.
Reliability check using Cronbach's Alpha

Reliability Statistics	
Cronbach's Alpha	N of Items
.769	24

It is widely known that the internal consistency of a survey can be evaluated using Cronbach's alpha, which checks if the items of the survey or a category of a survey are measuring the same concept (Cohen et al., 2017). As it can be noted form the above Table 1, an alpha of .769 is gained which is regarded as a generally acceptable to good value (Ahmed et al., 2024).

Table 2 displays the statistics for the items of the category related to factors that influence teachers' selection and implementation of assessment methods. As it is indicated in the table, Kurdish EFL university-level teachers agree that one of the most influential factors that affect their choice of assessment methods is *students' excessive reliance on AI tools,* with a mean value of 4.11 (SD= 0.879) and p=0.00 <0.05. This shows that teachers cannot properly employ assessment methods such as homework assignments, reports, projects, writing essays, portfolios and other sort of assessment methods which require students to write long blocks of essays. Moreover, factors such as

departmental guidelines (M = 3.69, SD=1.013, p-value = 0.00), having big number of students in the class (M = 3.52, SD=0.742 p-value 0.00), and the reliability issue of assessment methods (M = 3.65, SD=1.033, p-value = 0.00) suggest that teachers view these issues as influencing their choice of assessment methods. The teachers moderately agreed with the fact that the availability of resources influences their choice of assessment with mean value 3.21 (SD=0.810) and the difference was statistically significant (p-value 0.025), suggesting that this factor plays a meaningful role in shaping their choice of assessment methods, though the strength of the agreement was not high.

Table 2.
Factors influencing the selection and implementation of assessment methods

Items	N	Mean	Std. Dev.	t	P-value
Students' excessive reliance on AI tools influences teachers' selection of specific assessment methods.	75	4.11	.879	40.471	.000
Assessment methods are selected according to the standards set by departmental guidelines.	75	3.69	1.013	31.573	.000
Assessment methods are selected in terms of reliability.	75	3.65	1.033	30.624	.000
Classes of large numbers of students influence the selection of certain assessment methods.	75	3.52	.742	41.095	.000
The availability of resources, such as teaching facilities influences the choice of assessment methods.	75	3.21	.810	34.343	.025
Scoring criteria play a role in the selection of assessments methods.	75	3.12	.716	37.746	.151
Lack of proper in-service training makes it difficult to decide on the effective assessment methods.	75	3.11	.764	35.234	.230
Assessment methods that can be administered within the allocated time are used.	75	2.83	.742	32.989	.047

Table 2 also indicates that while time constraint might be considered, it is perceived as less influential as teachers have disagreed

that it affects their choice of assessment method with mean value 2.83 (SD=0.742), and p-value .047. This suggests that teachers do not employ assessment methods merely because they can be administered within a specific time frame. The rest of the factors: lack of proper in-service training and the scoring criteria with mean values: 3.11 (SD=0.764) and 3.12 (00.716) and p-values, .025 and .151 respectively, were not perceived as influential factors in the choice of assessment methods as their p-values were higher than the level of significance (0.05).

Table 3.
Challenges Encountered by Kurdish EFL University-Level Teachers in Assessment

ITEMS	N	Mean	Std. Devi.	t	P-value
Inconsistencies in students' performance make it hard to maintain fair assessment standards.	75	4.04	.861	40.628	.000
Limited access to technological resources impacts the quality of the assessment process.	75	4.04	.922	37.955	.000
Due to heavy reliance on AI tools, implementing assessment methods is challenging.	75	4.01	1.033	33.641	.000
Institutional policies and regulations restrict flexibility in assessment methods.	75	3.80	1.013	32.473	.000
Balancing between formative and summative assessments is a significant challenge.	75	3.79	.949	34.571	.000
Maintaining objectivity is quite challenging in grading.	75	3.68	1.164	27.372	.000
Assessing classes of large numbers of students is challenging.	75	3.67	.600	52.897	.000
The administrative workload related to assessments is overwhelming.	75	3.03	1.219	21.504	.005

The results in Table 3, overall, indicate a high level of teachers' agreement across the majority of the items in the survey, with seven out of eight items showing mean values ranging from 3.67 to 4.04. This indicates a strong consensus among Kurdish EFL university-level teachers with regard to the challenges they encounter in assessment. They collectively agree or strongly agree that inconsistencies in students' performance, limited access to technological resources, heavy reliance on AI tools, institutional polices restricting the flexibility of

assessment choice, balancing formative and summative assessment methods, maintaining objectivity in assessing, and having classes with large number of students with mean values 4.04 (SD=0.861, 4.04 (SD=0.922, 4.01 (SD=1.033), 3.80 (SD=1.013), 3.79 (SD=0.949), 3.68 (SD=1.164), and 3.67 (SD=0.600) respectively and their p-values collectively 0.00 which is <0.05. This underscores that such elevated mean values are not due to chance, but rather a genuine reflection and statistical robust perspectives of Kurdish EFL university-level teachers regarding the challenges they encounter in assessment.

An exception to the above statistics is the last item in Table (3) with a mean value of 3.03 (SD=1.219) and p-value of 0.05 suggesting a moderate level of agreement (leaning towards neutral). This indicates that teachers perceive assessment workload as not significant for they might believe that the influential factor of learning is the availability of valid and reliable assessment methods along with an effective basis of assessment.

Table 4.
Assessment Washback

ITEMS	N	Mean	Std. Dev.	t	P-value
Assessment methods are effective in identifying students' points of strength and weakness.	75	3.96	1.058	32.405	.000
Preparing for assessment motivates students to study hard and practice what they have learned.	75	3.83	.991	33.424	.000
Assessment methods encourage students' higher order thinking skills.	75	3.72	1.122	28.719	.000
The feedback from assessment methods positively impacts students' studying habits.	75	3.65	1.007	31.430	.000
Assessment makes students focus on authentic learning.	75	3.57	1.141	27.121	.000
Due to the critical nature of assessment students are heavily engaged with their learning tasks and hence study for longer times.	75	3.49	1.132	26.736	.000
Assessment drives students to memorize rather than understand concepts.	75	3.41	1.140	25.928	.002

Assessment positively impacts students' language abilities.	75	2.72	1.300	18.116	.066

Table 4 above presents teachers' perspectives of the washback impact of assessment on students' learning and studying habits. Except for the item that states that *assessment positively impacts students' language abilities* with a mean value 2.72 (SD=1.300) and a p-value 0.66 which is <0.05, the table indicates teachers' overall agreement with the majority of the items. Teachers collectively agree that assessment methods and practices conducted at university-level EFL context positively influence students learning and study habits in multiple ways, including identifying their strength and weaknesses, motivating them to study, fostering their higher-order thinking, shaping their study habits through feedback, promoting authentic learning, and increasing engagement, while also driving memorization. The mean values clustered within a narrow range (M =3.96 to 3.41) with all results statistically significant (p <0.05). It could be argued that the above statistics represent the real perspectives of Kurdish EFL teachers and therefore, together, these findings suggest that the pervasive influence of assessment in EFL context has positive impact on students' learning and studying habits. This could mainly be due to the fact that the assessment practices and methods employed are overall effective in supporting student learning and study habits.

DISCUSSION

The results reached in this study confirm that there are several factors that influence Kurdish university-level EFL teachers' choice and implementation of assessment. Among these factors are the departmental policies and guidelines, the availability of a big number of students in the class, the reliability of the assessment method in capturing a holistic picture of the potentials of the students, the availability of limited teaching and technological resources, and students' excessive reliance on AI tools in preparing their assignments, projects, and essays. The results of this study are in line with Cheng et al. (2007) and Latif and Wasim (2022) in that departmental guidelines

and standardized testing which are widely spread and more favored due to administrative requirements and accountability pressures, delimits the ability of teachers to employ modern and effective assessment methods. Moreover, this intensifies the challenges for teachers especially when there is a big number of students in the classroom, as this has been found to be another factor affecting teachers' choice and implementation of assessment methods. The finding is in harmony with Latif and Wasim (2022) and Hussain et al. (2019) as, in their studies, they found out that having classes with big number of students limits teachers' ability to employ diverse assessment methods. As a result, this may in turn force the teachers to rely less on valid, robust, and reliable assessment methods in favor of those promoting surface learning and drive students to study only for the test. Students' excessive reliance on AI tools in assessment has been found out to be one of the most crucial factors that influence teachers' choice and implementation of assessment methods especially those authentic ones. In this regard, this finding aligns with the results of many studies investigating the impact of AI on assessment, among these is the study by Tanberkan et al. (2024) who also found out that those students lacking AI ethical literacy may choose to use it in an unethical manner such as provision of ready-made assignments, projects, and other educational requirements.

Regarding the challenges teachers encounter in assessment, the findings in the current study reveal that Kurdish university-level EFL teachers face multiple, interrelated challenges. Among the most frequently reported challenges are students' excessive dependence on AI-driven websites in preparing their assignments and projects, the lack of adequate technological infrastructure, varying levels of student performance, overcrowded classrooms, and the constant struggle to balance formative and summative assessment. Furthermore, maintaining objectivity in grading and scoring remains a persistent concern. These challenges align with those identified in earlier research. For instance, Tanberkan, et al. (2024) found out that teachers tend to question the originality of student work due to the widespread

use of generative AI, which undermines the integrity of assessments. However, due to a lack of applications and software that check the originality of the work, many teachers depend on those which are available on Google and are free. Similarly, a study by Brown and Abeywickrama (2019) highlighted that insufficient teaching facilities and devices compromise the fair implementation of assessment, especially the technology-integrated assessments. In line with this, studies such as Hussain et al. (2019) and Tsagari and Vogt, (2017) have found out that large class sizes significantly impact teachers' choice of assessments and the feasibility of the provision of tailored feedback, which is seen as a critical element in assessment. Also, teachers struggle to be objective in grading and scoring students' answers when they are faced with imbalanced levels of students' academic abilities (Malone, 2022). These findings are in harmony with the results of the current study, highlighting that assessment challenges are both universal and multifaceted.

As it could be noted in the results section and overall positive washback effects of assessment have been noticed. This is reported to generally encourage the provision of positive feedback that enhances students' studying habits. Also, positive washback of assessment was eminent in motivating students to study hard, pinpointing their weak and strong aspects of language learning, encouraging their higher-order thinking skills, and directing students' focus more to authentic types of learning. These results are in line with the results from Albondoq (2023) reporting that there is positive washback in assisting students to learn, driving them to become motivated, and developing their language skills. A negative washback effect of assessment was reported to being ineffective in developing their overall language abilities. This result is in line with the results from Omar (2020), who found out that assessments, especially tests, had negative washback effect on teaching and learning.

CONCLUSION

The current study concludes that there are a number of factors that hinder University-level EFL teachers' selection and implementation of assessment methods, such as strict departmental guidelines, students' reliance on AI tools, the availability of a big number of students in the classes, the lack of teaching facilities, and the reliability issue of assessment method. It also concludes that teachers encounter a range of challenges when implementing assessment including: students' reliance on AI tools, the availability of large number of students in the classrooms, students' proficiency level, teachers' limited access to technological resources especially software or applications that detect AI-written texts, abiding by strict institutional guidelines, balancing formative and summative types of assessment, as well as maintaining objectivity in grading. In terms of washback, it is concluded that, overall the assessment methods and practices conducted in the EFL context yield effective positive washback in that they foster motivation, higher-order thinking, engagement, authentic learning, enhance study habits, and pinpoint students' strengths and weaknesses in language learning. At the same time, the finding that assessment is ineffective in developing students' language abilities underscores the need for carefully designed assessment methods that minimize negative consequences.

Based on the above conclusions, it is recommended that EFL departments at the public universities in Kurdistan Region of Iraq reform their departmental guidelines and provide teachers with more autonomy in choosing assessment methods based on their reliability and students' needs. Also, institutions and universities must provide teachers with valid and reliable AI detection software so that students would not heavily depend on AI tools in their assessments. Moreover, it is recommended that universities or EFL departments be given the freedom to decide about the number of students admitted to the department. To better support fair and innovative assessment, the ministry of higher education and scientific research is recommended to provide all teachers with adequate access to modern technological

devices and teaching facilities. This would help them use diverse methods of authentic assessments. Finally, universities should work on providing ongoing professional development programs for the staff of the EFL departments to improve their assessment literacy.

DECLARATION OF AI AND AI-ASSISTED TECHNOLOGIES

The authors declare that they have used ChatGPT solely for the purpose of improving the language and readability of the manuscript. They have not used any AI-assisted technologies for designing, analyzing, and interpreting the data of this research and that all the contributions remain the sole work of the authors. They assume full responsibility for the publication of the content of the paper. In line with the guidelines and polices of the journal concerning the use of AI-assisted technologies, this declaration is prepared.

REFERENCES

Ade, J. H. (2021). Challenges in the implementation of the Bologna process. *Polytechnic Journal, 11*(2), Article 13. https://doi.org/10.25156/ptj.v11n2y2021.pp79-86

Ahmed, A., Pereira, L., & Jane, K. (2024). *Mixed methods research: Combining both qualitative and quantitative approaches*. https://www.researchgate.net/publication/384402328

Albondoq, I. S. A. (2023). Washback effects on English language learning and teaching at Aden university. *Journal of English Studies in Arabia Felix*, 2(2), 36–45. https://doi.org/10.56540/jesaf.v2i2.66

Allen, M.J. (2004). *Assessing academic programs in higher education*. San Francisco: Jossey-Bass.

Arthur, F. and Partey, P. (2023). Washback effects of high-stakes test: Perspectives of economics teachers and students. *Asian Journal of Education and Social Studies*, 40(1), 52-67. https://doi.org/10.9734/ajess/2023/v40i1866

Bachman, L. F., & Damböck, B. (2018). *Language assessment for classroom teachers.* Oxford University Press.

Bailey, K. M., & Brown, J. D. (2022). *Language testing courses: What are they?* Language Testing, 39(1), 153–174. https://doi.org/10.1177/02655322211026876

Black, P., & Wiliam, D. (2018). Classroom assessment and pedagogy. *Assessment in Education*, 25(6), 551–575. https://doi.org/10.1080/0969594X.2018.1441807

Bouckaert, M. (2023). *The assessment of students' creative and critical thinking skills in higher education across OECD countries: A review of policies and related practices* (OECD Education Working Papers No. 293). Organisation for Economic Co-operation and Development. https://doi.org/10.1787/35dbd439-en

Brookhart, S. M. (2017). *How to use grading to improve learning*. ASCD. https://lccn.loc.gov/2017014773

Brookhart, S. M., & Nitko, A. J. (2019). *Educational assessment of students* (8th ed.). Boston, MA: Pearson.

Brown, D. H. (1990). *Language assessment: Principles and classroom practices*. London.

Brown, J., & Abeywickrama, P. (2019). *Language assessment: Principles and classroom practices* (3rd ed.). Pearson Education.

Chappuis, J. and Stiggins, R. (2017) *Introduction to Students-Involved Assessment for Learning* (7th ed.). Pearson Education, Inc., Boston.

Cheng, L., & Curtis, A. (Eds.). (2010). English Language Assessment and the Chinese Learner (1st ed.). Routledge. https://doi.org/10.4324/9780203873045

Cheng, L. (2005). *Changing language teaching through language testing*. Cambridge University Press.

Cheng, L., Rogers, W. T., & Wang, X. (2007). Assessment purposes and procedures in ESL/EFL classrooms. *Assessment & Evaluation in Higher Education*, *33*(1), 9–32. https://doi.org/10.1080/02602930601122555

Chhetri, D., & Khanal, B. (2024). A pilot study approach to assessing the reliability and validity of relevancy and efficacy survey

scale. *Janabhawana Research Journal*, 3(1), 35-49. https://doi.org/10.3126/jrj.v3i1.68384

Cohen, L., Manion, L., & Morrison, K. (2017). *Research Methods in Education*. New York, NY: Routledge. https://doi.org/10.4324/9781315456539

Creswell, J.W. and Creswell, J.D. (2018) *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage, Los Angeles.

Darling-Hammond, L., Wilhoit, G., & Pittenger, L. (2022). *Assessment for a changing world: Reimagining student learning for the future*. Learning Policy Institute. https://learningpolicyinstitute.org/product/assessment-changing-world-report

Deneen, C.C., Fulmer, G., Brown, G., Tan, K., Leong, W.S., & Tay, H.Y. (2019). Value, practice and proficiency: Teachers' complex relationship with assessment for learning. *Teaching and Teacher Education*, 80, 39 47. https://doi.org/10.1016/j.tate.2018.12.022 .

Erlinawati, E., & Muslimah, M. (2021). Test validity and reliability in learning evaluation. *Bulletin of Community Engagement, 1*(1), 26–31.https://www.attractivejournal.com/index.php/bce/article/view/1

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education* (8th ed.). New York: Mc Graw HIll.

Fulcher, G. (2020). *Practical language testing*. Routledge.

Fulmer, G. W., Lee, I. C. H., & Tan, K. H. K. (2015). Multi-level model of contextual factors and teachers' assessment practices: An integrative review of research. *Assessment in Education: Principles, Policy & Practice*, 22(4), 475–494. https://doi.org/10.1080/0969594X.2015.1017445 .

Giraldo, F. (2021). Language assessment literacy and teachers' professional development. *ELT Journal*, 75(3), 356–367. https://doi.org/10.1093/elt/ccab020

Great Schools Partnership. (2015). *Assessment*. The Glossary of Education Reform. https://www.edglossary.org/assessment/

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81-112. https://doi.org/10.3102/003465430298487

Hamp-Lyons, L. (2017). *Assessing English for academic purposes*. Cambridge University Press.

Hussain, S., Shaheen, N., Ahmad, N., & Islam, S. U. (2019). Teachers' classroom assessment practices: Challenges and opportunities to classroom teachers in Pakistan. *The Dialogue*, *14*(1), 87–102. https://www.qurtuba.edu.pk/thedialogue/The%20Dialogue/14_1/Sajjad%20Hussain.pdf

Kang, E., & Hwang, H.-J. (2023). The importance of anonymity and confidentiality for conducting survey research. *Journal of Research in Public Education, 4*(1), Article 1. https://doi.org/10.15722/jrpe.4.1.202303.1

Kozma, R. B. (2003). *ICT and educational change: A global phenomenon*. In R. B. Kozma (Ed.), Technology, innovation, and educational change: A global perspective (pp. 1–18). Eugene, OR: International Society for Technology in Education.

Kunnan, A. J. (2018). *Evaluating language assessments.* Routledge.

Latif, M. W., & Wasim, A. (2022). Teacher beliefs, personal theories and conceptions of assessment literacy—a tertiary EFL perspective. *Language Testing in Asia, 12*(1), Article 25. https://doi.org/10.1186/s40468-022-00158-5

Lieberman, J. D., Krauss, D. A., Heen, M., & Sakiyama, M. (2016). The good, the bad, and the ugly: Professional perceptions of jury decision-making research practices. *Behavioral Sciences & the Law*, 34(4), 495–514. https://doi.org/10.1002/bsl.2246

Ma, M. (2023). Correction: Exploring learning-oriented assessment in EAP writing classrooms: teacher and student perspectives. *Language Testing in Asia,* 13, 47. https://doi.org/10.1186/s40468-023-00264-y

Ma, M., & Bui, G. (2021). Chinese secondary school teachers' conceptions of L2 assessment: A mixed-methods study. *Studies*

in Second Language Learning and Teaching, 11(3), 445–472. https://doi.org/10.14746/ssllt.2021.11.3.7

Malone, M. E. (2022). Training for rater reliability. *Language Testing*, 39(2), 245–267. https://doi.org/10.1177/02655322211034567

Mertens, D. M. (2019). *Research and evaluation in education and psychology: Integrating diversity with quantitative, qualitative, and mixed methods*. Sage Publications.

Omar, J. A. (2020). Negative washback of the 12th Grade English Language Exam on teaching and learning in the Kurdistan Region of Iraq. *Journal of University of Raparin, 7*(3), 672–690. https://doi.org/10.26750/Vol(7).No(3).paper29

Perkins, M., Gezgin, U. B., & Roe, J. (2023). AI and plagiarism in EFL writing. *Journal of Academic Ethics*, 21(1), 89–105. https://doi.org/10.1007/s10805-022-09451-9

Popham, W. J. (2018). *Classroom assessment: What teachers need to know* (8th ed.). Pearson.

Shadiev, R., & Hwang, W.-Y. (2020). Review of technology-assisted language assessment. *Computer Assisted Language Learning*, 33(8), 1–25. https://doi.org/10.1080/09588221.2020.1770292

Sherif, V. (2018). Evaluating Preexisting Qualitative Research Data for Secondary Analysis. *Forum Qualitative Sozialforschung Forum: Qualitative Social Research*, *19*(2). https://doi.org/10.17169/fqs-19.2.2821

Tanberkan, H. Özer, M. & Gelbal, S. (2024). Impact of artificial intelligence on assessment and evaluation approaches in education. *International Journal of Educational Studies and Policy*, 5(2), 139 152. https://doi.org/10.5281/zenodo.14016103

Tsagari, D., & Vogt, K. (2017). Assessment literacy of foreign language teachers around Europe: Research, challenges and future prospects. *Papers in Language Testing and Assessment*, 6(1), 41–63. https://arts.unimelb.edu.au/__data/assets/pdf_file/0010/2349928/6_1_SI3TsagariVogt.pdf

Wächter, B. (2004). The Bologna process: Developments and prospects. *European Journal of Education, 39*(3), 265–273.

Warschauer, M. (2020). AI and the future of language learning. *CALICO Journal*, 37(2), 1–12. https://doi.org/10.1558/cj.41123

Weigle, S. C. (2020). *Assessing writing* (2nd ed.). Cambridge University Press.

Williams-McBean, C. (2025). Factors influencing teachers' choice and use of assessment. *International Journal of Studies in Education and Science* (IJSES), 6(2), 212-239. https://doi.org/10.46328/ijses.129

Wilson, M. (2024). Finding the right grain-size for measurement in the classroom. *Journal of Educational and Behavioral Statistics*, 49(1), 3-31. https://doi.org/10.3102/10769986231159006wilsom

Yan, Z., Chiu, M. M., & Cheng, E. C. K. (2022). Predicting teachers' formative assessment practices: Teacher personal and contextual factors. *Teaching and Teacher Education*, 114. Retrieved from https://doi.org/10.1016/j.tate.2022.103718

Yin, M. (2010). Understanding classroom language assessment through teacher thinking research. *Language Assessment Quarterly*, *7*(2), 175–194. https://doi.org/10.1080/15434300903447736

APPENDIX Teachers' Survey

Teachers' Survey

Dear Teachers,

You are kindly invited to participate in the following survey, which is expected to take no more than 5 minutes of your time. Your cooperation is highly appreciated, and your thoughtful responses are essential to the success of this study. The study focuses on teachers' perceptions of the factors influencing their assessment choice, the challenges they face, and the washback effect of assessment.

Please note that your participation is entirely voluntary, and your responses will be treated with strict confidentiality. The information gathered through this survey will be used solely for academic research purposes and will remain anonymous. You are free to withdraw from the survey at any point without any negative consequences. Your insights are of great value and will significantly contribute to the depth and quality of this study.

Thank you for your time and contribution.

1. Demographic Information
   Gender: Male ☐ Female ☐
   Qualification: PhD ☐ Master ☐
2. Years of experience:
   Less than 5 years ☐ 5-10 years ☐ 11-20 years ☐ more than 20 years ☐
3. Academic title:
   Professor ☐ Assistant Professor ☐ Lecturer ☐ Assistant Lecturer ☐
4. University : University of Zakho ☐, University of Duhok ☐, University of SalahAldin ☐, University of Sulaimani ☐, University of Halabja ☐, University of Soran ☐, University of Garmian ☐, University of Charmo ☐, University of Koya ☐, University of Raparin ☐, and University of Acre ☐

FACTORS INFLUENCING THE SELECTION AND IMPLEMENTATION OF ASSESSMENT METHODS						
No.	Items	SA	A	N	D	SD
1	Assessment methods are selected according to the standards set by departmental guidelines.					
2	Assessment methods that can be administered within the allocated time are used.					
3	Classes of large numbers of students influence the selection of certain assessment methods.					
4	Assessment methods are selected in terms of reliability.					
5	The availability of resources, such as teaching facilities influences the choice of assessment methods.					

6	Lack of proper in-service training makes it difficult to decide on the effective assessment methods.					
7	Scoring criteria play a role in the selection of assessments methods.					
8	Students' excessive reliance on AI tools influences teachers' selection of specific assessment methods.					
CHALLENGES FACED BY EFL TEACHERS IN THE ASSESSMENT PROCESS						
No.	**Items**	**SA**	**A**	**N**	**D**	**SD**
1	Due to heavy reliance on AI tools, implementing assessment methods is challenging.					
2	Assessing classes of large numbers of students is challenging.					
3	Inconsistencies in students' performance make it hard to maintain fair assessment standards.					
4	Limited access to technological resources impacts the quality of the assessment process.					
5	Institutional policies and regulations restrict flexibility in assessment methods.					
6	Balancing between formative and summative assessments is a significant challenge.					
7	Maintaining objectivity is quite challenging in grading.					
8	The administrative workload related to assessments is overwhelming.					
WASHBACK OF ASSESSMENT						
No.	**Items**	**SA**	**A**	**N**	**D**	**SD**
1	Assessment positively impacts students' language abilities.					
2	The feedback from assessment methods impact students' studying habits.					
3	Preparing for assessment motivates students to study hard and practice what they have learned.					
4	Assessment methods are effective in identifying students' points of strength and weakness.					
5	Assessment methods encourage students' higher-order thinking skills.					
6	Due to the critical nature of assessment, students are heavily engaged with their learning tasks and hence study for longer times.					
7	Assessment makes teaching focus on authentic learning.					
8	Assessment drives students to memorize rather than understand concepts.					

Thank you very much for your participation.