Open Access

SHIFTING PEDAGOGIES: THE ROLE OF EXAM'S FORMAT AND STAKE ON SECONDARY SCHOOL LANGUAGE TEACHERS' INSTRUCTIONAL DECISIONS

AZIZI ALIAS
NATIONAL UNIVERSITY OF MALAYSIA, BANGI, SELANGOR, MALAYSIA
NATIONAL UNIVERSITY OF MALAYSIA, NAMGI, SELANGOR, MALAYSIA

KAMISAH OSMAN
NATIONAL UNIVERSITY OF MALAYSIA, NAMGI, SELANGOR, MALAYSIA

ABSTRACT:

High-stakes language examinations often shape not only student behavior but also the instructional decisions of language teachers. This phenomenon, known as washback, plays a critical role in how teaching practices align with assessment demands. This study intents to explore the nature and extent of washback effects on language teachers' instructional strategies due to exam's format and its stakes. A quantitative approach was employed, involving the development of washback survey, administered on 370 language teachers and multiple linear regression analysis was assisted to determine the roles of exam's format and stakes on teacher exam-oriented practice. The model was statistically significant, $F(2, 367) = 170.00$, $p < .001$, explaining 48.2% of the variance in assessment decisions. Both exam's format ($\beta = 0.583$, $p < .001$) and stakes ($\beta = 0.319$, $p < .001$) were significant predictors, with exam's format exerting a stronger influence. Findings revealed that teachers frequently adapt their teaching methods and materials to reflect exam formats and perceived expectations. The study contributes to a deeper apprehension of how assessment policy influences classroom practice and the substantial impact of test design and perceived stakes on teacher behavior in language assessment contexts, with implications for assessment literacy, policy development, and test use in high-stakes environments. Future research should further investigate contextual variables that mediate the intensity and direction of washback in diverse educational settings.

Keywords: Washback, exam-oriented, language teachers, language assessment

INTRODUCTION:

The demand for effective communication skills has become paramount as globalization increases. In a globalized economy, the proficiency in language significantly impacts career opportunities through social mobility and enhanced employability. Communication skills are crucial for employability with studies indicating that graduates with strong communicative skills have higher employment rates (Grover & Pandya, 2024). Since individuals may be judged based on their language skills, impacting their career advancement opportunities (Greer et al., 2024). The ability to navigate multilingual environments enhances professional identities and opens up new career pathways (Goble, 2024). More than 80% of students passed English in (Malaysian Education Certificate) SPM but most were not able to speak English very well (Mohd Nazri Latiff & Rosnah, 2018). It shows that the students have the capabilities to perform well in exam but not mastering the language itself. High-stakes exams, such as for university entry, can lead to a misalignment between curriculum goals and teaching practices, ensuing in a focus on test preparation relatively than comprehensive language learning (Allen et al., 2024). Thus, do existing assessment method and high-stakes examination validly measure students' communications as demanded by today's economic demands.

Examination in education settings promote change in attitude and behavioral change towards language learning and the exam. Teachers tend to do practical and more authentic activities but remain teacher-centered approach (Cheng, 1997). High-stakes exams cause students and teachers to focus on test-taking strategies rather than broader language skills (Ali, 2024). Exam-oriented amongst language teachers leading to exam preparation focusing on teaching methodology, learning strategies and choices of materials (Aftab, Qureshi & William, 2014). Exam-oriented making language learning more on getting good score in exam but not on the application of the language (Andrews, Fullilove & Wang, 2002). Thus, the result of the exam affects how teachers teach. The teachers will then adjust their teaching accordingly to fit the examination and the students who receive the modified teaching to take the examination which will generate another round of result for teachers to modify their teaching again (Chan, 2020).

High-stakes examinations has frequently strived a strong impact on teaching and learning in educational environment (Chou 2017). Testing continues to be widely used because it is often seen as the main way to bring changes to the education system without altering other parts, like teacher training or the curriculum. However, this view is considered naive and overly simplistic (Cheng, Sun & Ma, 2015). Most researchers studying the effects of exams or the existence of washback either in general or in the area of language education focus on the psychometric aspects of the exams without providing empirical evidence regarding the effects of the exams themselves (Alla Baksh et al., 2019). There are several studies that show that guidance on negative evaluation is directly proportional to improved achievement in language tests (Mohd Nazri Latiff & Rosnah, 2018). Despite the abolition of national exam, teachers continue using past exam papers, indicating residual washback effects (Auladi, 2024).

Students show a preference for studying language components that frequently appear in high-stakes exams and allocate more time to practicing exam-related content (Desalegn, Disassa & Kitila, 2023). Language teachers feel guilty and frustrated because unable to make content more interesting to students and not able to improve scores of their student (Alderson & Hamp-Lyons, 1996). Due to the significant importance of university entrance exam, high school graduates often enroll in exam preparation classes at cram schools (Allen, 2016). Tsang (2017) revealed that there are multiple washback effects brought by the examination to the students which included the needs to participate in rigorous exam practice sessions and supplementary tutoring classes. This shows that students are more focused on getting good results in exams than improving their language skills and give teachers few options to prepare students for the exams than to teach authentically.

Examinations such as CESOL influenced the selection of teaching resources used by English teachers and students during classroom activities (Hoa, 2020). In brief, findings from the analysis revealed that English teachers and students made use of multiple supplementary materials, including commercial texts, academic journals, books, and exam-oriented writing exercises seems to be prominent. Exam-oriented culture persists, with teachers still prioritizing test-focused exercises although teachers perceive School Based Assessment (SBA) as more practical, but still rely on predictable test formats (Auladi, 2024). Studies show that the choice of exam format can influence student performance, with certain formats yielding better results in specific contexts (Bordel et al., 2023). Different item types in test can influence students' understanding and performance, prompting teachers to choose materials that mirror exam formats (DeCarlo, 2024). Although SBA is considered beneficial, teachers still depend on exam-based preparation, as the pressure of exams reinforces the use of predictable formats and test-aligned materials.

WASHBACK IN LANGUAGE TESTING

Examination or test have the capability to alter or influence the process of teaching and learning or commonly known as washback (Bailey, 1996). Exams results specifically from high-stakes test affect the exam outcomes and form the behavior of the stakeholders (Tsagari and Cheng, 2017). The impact of assessment such like test occurs not only at the end of the instructional process but prior to the test itself and throughout the preparation stage (Brown, 2004). Messick (1996) noticeably explains that washback intentionally or unintentionally has short- or long-term repercussion of score evaluation and use for individuals, institutions, and society. Washback is a highly complex phenomenon and construct and have been theorized in multiple ways (Alderson & Wall, 1993; Rea-Dickins & Scott, 2007; Green, 2007; Spratt, 2005; Watanabe, 2004). Figure 1 shows the aspects of washback studies encompassing its dimension, the 3Ps model and factors impacting washback process (Watanabe, 2004).
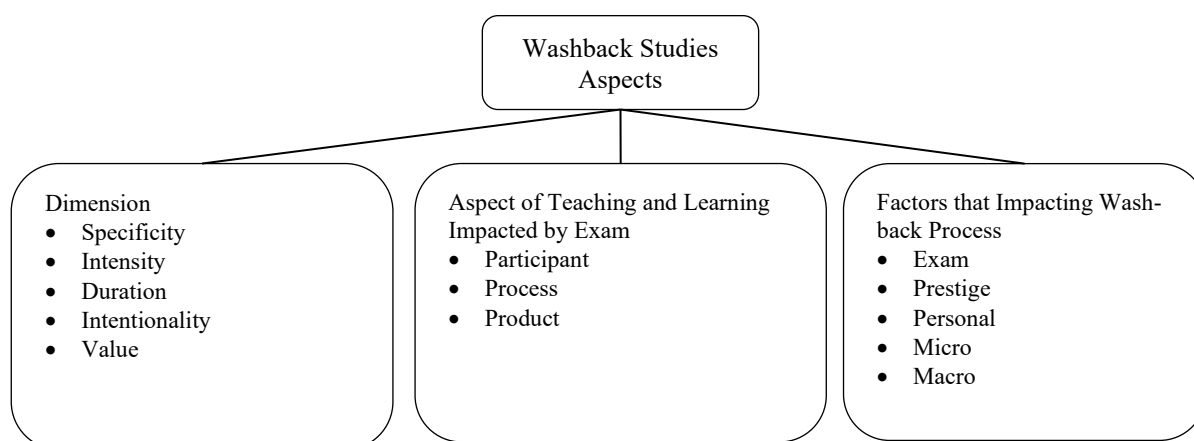

Figure 1: Aspects in Washback Studies
Source: Watanabe 2004

Traditional large-scale multiple-choice exams are often criticized for their negative washback on curriculum design, teaching methods, and classroom activities (Chou, 2017). Empirical research has further demonstrated that testing affects teaching content, instructional approaches, assessment practices, and overall teaching styles (Cheng, Sun, & Ma, 2015). Test-oriented pressures cause teachers and students to perceive exam preparation as separate from authentic English learning (Chak,2024). Memorization was the dominant strategy, with students prioritizing rote learning over language skills development. (Dawadi, 2021). Grammar and vocabulary exercises dominate students' learning habits, influenced by exam content. (Desalegn, Disassa, & Kitila, 2023). English teachers in the present study did not teach listening, but adopted a so-called 'doing' or 'testing' approach (Chou, 2019). Listening and speaking skills remain underassessed, despite curriculum recommendations (Auladi, 2024). Teachers are more focus on skills such as listening and speaking but the content of teaching is based around the content of the exam (Cheng, 1997). Teachers may resort to teaching methods that prioritize test performance over genuine language acquisition, as evidenced by mixed washback effects observed (Wang et al., 2024).

Increasing teachers' pedagogical knowledge of teaching listening can be achieved when teachers are allowed to examine their own current beliefs and practices (Chou, 2019). Teachers' knowledge of the SWU-SET influenced their lesson planning, leading to more test-focused instruction (Athiworakun & Adunyarittigun, 2024). Negative effect of exam on language teaching is when the knowledge and skill were narrowed to the format and content of the exam not the learning of the language as subjected in the objective (Aftab, Qureshi & William, 2014). Teachers are not to be blame because of the selection of materials based on examination and they prefer materials that resemble to the examinations due to student's anxiety to succeed in the exam (Alderson & Hamp-Lyons 1996). Prior studies showed that a more test-oriented pedagogy limiting opportunities for interaction and instruction emphasis more on individual activities than group or pair work, compared to a more general English class (Clark & Yu, 2022).Teachers are reluctant to fully embrace communicative teaching, as they prioritize test preparation (Auladi, 2024). As suggested by Chan in the Figure 2. below, teacher's decision were based on test results and modify their teaching according to what is demanded or the format of the exam.
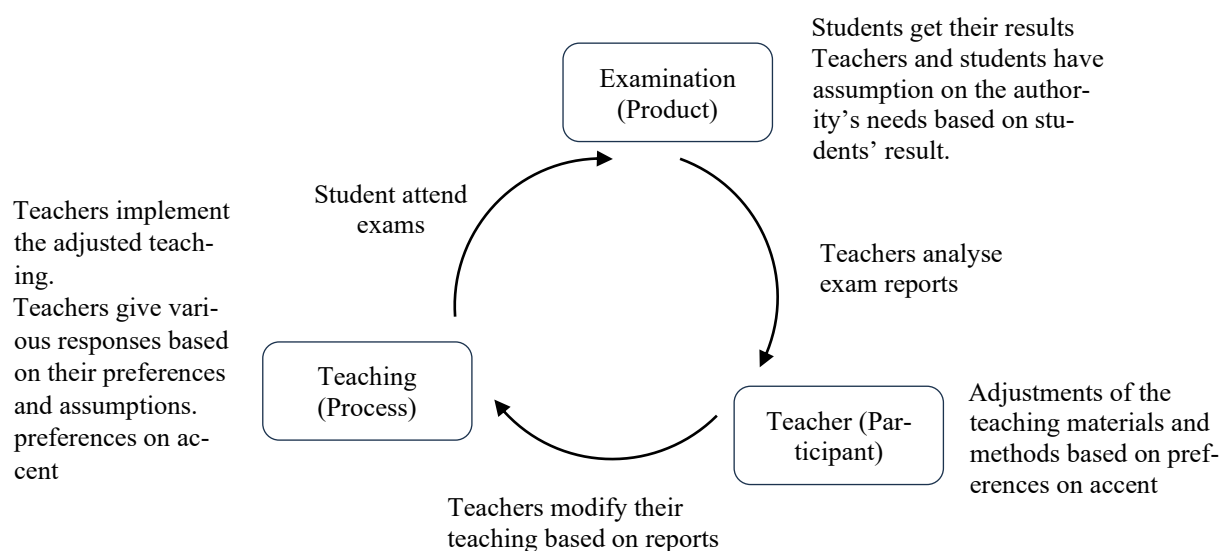

Figure 2: Washback Cycle for Teachers (Chan, 2020)

Students expect that the materials they learn will aligned the exam format and when this is not the case, they may feel ineligible leading to discouragement (Abate & Mishore, 2024). Grammar-focused exam tasks discourage students from engaging with English in a natural, communicative way (Chak, 2024). Language skills that have high weighting, such as reading skills, grammar and vocabulary, as well as examination-taking techniques, are the main activities in the classroom (Allen, 2016). Students' performance in the exam are getting better when the teachers are more familiar with the exam format and its requirement (Andrews, Fullilove & Wang 2002). High-stakes testing can exacerbate anxiety among learners, potentially hindering their performance and overall learning experience (Zarei et al., 2024). Students may prioritize test-specific skills over broader language competencies, which can limit their overall language development (Ye et al., 2024). The pressure associated with these exams can drive students to adopt more intensive study habits, but it may also limit the diversity of their learning approaches.

The researchers challenged the bureaucrats' belief that both tests had reached their objectives with no need for teacher training or curricular revision. The study concludes that washback changes with time because of factors such as language status and test use in society (Cheng, Sun & Ma, 2015). The framework does not assume a testing-teaching causal relationship but instead, the testing-teaching relationship is in the wider socio-political context in which other social factors including those related to the examination influence teaching (Ali & Hamid 2020). Without understanding washback, opportunities for professional growth and curriculum alignment may be missed, limiting teachers' ability to adopt effective, proficiency-oriented practices (Vyn, 2024). Historical perspectives show that neglecting washback can perpetuate a fragmented assessment landscape, as seen in Singapore's evolution from colonial testing practices to centralized governance (Zakaria & Aryadoust, 2024).

Most previous studies on washback have focused on qualitative findings (Dawadi, 2021; Hennessy & Nogueira, 2024; Mahmood, Mobeen & Abbas, 2020).) and lack of empirical evidence (Ahmed & Sikder, 2023; Baharuddin et al., 2020; Caner & Bayhan 2020). Therefore, this study offers findings that can be generalized in the context of factors that mediate washback on teachers' decisions. It directly contributes to the body of knowledge related to the theory of teacher decisions that needs to be considered before a new assessment policy is introduced. This study aims to determine the exam factors or stakes that more influence teacher exam-oriented practice to prove the following hypothesis;

(H₀): The overall regression model does not explain a significant amount of variance in Y.

$H_0: R^2 = 0$

Efforts in identifying and empirically testing potential mediating factors in this study contributes to theoretical understanding of washback and how these factors influence teacher's classroom decisions. By explicating these pathways, the research deepens apprehension into the complex process through which washback operates. Additionally, the study establishes and validates a survey respectively designed to measure factors that mediate washback effect. The validated survey provides an attested and practical tool for researcher and stakeholders to pursue these mediating factors in diverse educational context. The study underpins more informed educational policy and classroom decision. The findings can enlighten interventions aimed at alleviate negative washback and enhancing positive impacts of assessment in classroom instruction.

METHOD

A mixed-methods sequential exploratory design was employed in this study starting with a qualitative approach for survey development and continuing with a quantitative phase for survey validation and analysis. The research was designed to investigate whether exam's format and exam's stakes serve as predictors of teachers' exam-oriented practice. To construct a contextually valid washback survey, qualitative data were first gathered through semi-structured interviews with nine experts in language education, assessment, and curriculum. Thematic analysis of the interview transcripts guided the development of initial survey items. The experts were selected based on their experience in language teaching, assessment policy, and involvement in the national SPM examination system.

To ensure the content validity of the survey, the Fuzzy Delphi Method (FDM) was employed. 15 panel of experts evaluated the relevance, clarity, and necessity of the proposed items. FDM allowed for systematic consensus building by calculating threshold values ($d \leq 0.2$), expert agreement percentages ($\geq 75\%$), and defuzzification values (≥ 0.5), leading to the refinement and validation of the survey content (Aidil et al., 2023; Manakandan et al., 2017; Ramalingam et al., 2019). 30 out 43 items remain and was piloted to 105 language teachers in secondary school and further elimination of items due to miss fitting items leaving only 25 item using Rasch Measurement Model. Analysis indicated acceptable item fit, unidimensionality, and high person reliability indices, affirming the survey's robustness and consistency in measuring washback-related constructs using Winsteps.

Following validation, the finalized survey was distributed to a larger sample of 370 language teachers across Malaysia. All participants were currently teaching language that are assessed in the national SPM high-stakes examination, thereby ensuring relevance to the study's focus on washback effects in a high-pressure assessment environment. Quantitative data from the main survey were analyzed using multiple linear regression analysis to explore the predictive roles of two independent variables which are exam format and exam stakes, on the dependent variable teachers' exam-oriented instructional practices. The analysis was conducted using JAMOVI, and assumptions of normality, linearity, and multicollinearity were tested and met. Ethical approval for this study was obtained from the relevant institutional review board prior to data collection. Several ethical principles were adhered to throughout the research process to ensure the protection and rights of all participants.

RESULTS

A Rassch Measurement Model was applied to examine the psychometric properties of the instruments using Winsteps. The analysis included fit statistics, unidimensional and assessment of local dependence. Using infit dan outfit mean square statistics (MNSQ), the item fit was assessed with acceptable values ranging 0.50 to 1.50 (Table 1) were considered productive for Rasch Model assessment (Linacre, 2015) and the value of Zstd can be flagged if the infit and outfit values are accepted (Linacre, 2005). The Median Absolute Deviation of Adjusted Q3 (MADaQ3) residuals was calculated the value below 0.20 (Table 2), suggest the assumption of local independence is confirmed (Christensen, Makransky, & Horton, 2017). Appropriate item discrimination referring to Point-Measure Correlation indicate more than 0.20 and dimensionality was further determined using Principal Components Analysis (PCA) of residuals showed a first contrast eigenvalue below 2.0 (Table 3) was representative of unidimensionality and suggesting no substantial secondary dimension (Bond & Fox, 2015).

Table 1: Analysis of Item Fit

ITEM	MODEL		INFIT		OUTFIT		PTMEASUR-AL	
	MEASURE	S.E.	MNSQ	ZSTD	MNSQ	ZSTD	CORR.	EXP.
P1	-0.38	0.1	1.20	2.38	1.43	3.61	0.60	0.65
P2	-0.91	0.11	1.11	1.3	1.29	2.11	0.58	0.60
P3	-0.50	0.1	1.27	3.13	1.38	3.10	0.57	0.64
P4	-0.27	0.1	0.81	-2.46	0.76	-2.44	0.70	0.65
P5	-0.13	0.1	0.84	-2.04	0.82	-1.90	0.70	0.66
Q1	-0.20	0.1	1.02	0.32	1.11	1.04	0.66	0.66
Q2	-0.51	0.1	1.04	0.52	1.12	1.07	0.62	0.64
Q3	-0.78	0.11	0.93	-0.82	0.98	-0.08	0.62	0.61
Q4	-1.12	0.11	0.88	-1.59	0.82	-1.32	0.61	0.59
Q5	-1.09	0.11	1.07	0.87	1.36	2.40	0.55	0.59
Q6	-0.05	0.1	1.00	0.06	0.96	-0.43	0.66	0.67
Q7	-0.95	0.11	0.94	-0.7	0.92	-0.57	0.61	0.60
Q8	-0.22	0.1	1.28	3.19	1.32	2.87	0.60	0.66
R1	0.59	0.1	0.83	-2.15	0.85	-1.73	0.75	0.71
R2	0.75	0.09	0.82	-2.29	0.85	-1.84	0.74	0.72
R3	1.30	0.09	0.98	-0.26	1.00	0.06	0.75	0.74
R4	1.30	0.09	0.96	-0.51	0.98	-0.20	0.76	0.74
R5	0.94	0.09	0.79	-2.77	0.77	-2.91	0.78	0.73
R6	0.36	0.10	0.85	-1.94	0.79	-2.51	0.74	0.69
R7	0.35	0.10	0.90	-1.17	0.86	-1.58	0.73	0.69
R8	1.58	0.09	1.38	4.27	1.48	4.25	0.68	0.75
R9	0.34	0.1	1.05	0.68	1.23	2.43	0.65	0.69
R10	-0.71	0.11	0.90	-1.24	0.80	-1.76	0.65	0.62
R11	-0.26	0.10	1.00	0.06	1.02	0.22	0.67	0.65
R12	0.58	0.10	0.93	-0.79	1.03	0.35	0.71	0.71

Table 2: Model Fit and Local Dependency Test						
		Person Reliability		MADaQ3		p
Scale		0.924		0.179		< .001

Table 3: Unidimensionality Analysis

Sub Scale	Raw variance explained by measures % Observed	Raw variance explained by measures % Expected	Unexplained variance in 1st contrast Eigenvalue	Disattenuated correlation of cluster 1-3
Exam's Format (5 Items)	51.4	51.2	1.6	>0.9
Exam's Stakes (8 items)	54.4	54.3	2.0	>0.9
Exam-Oriented Practice (12 items)	60.3	59.9	1.9	>0.9

The examination to which predictors of Exam's Format or Exam's Stakes that predicts Exam-Oriented Practice a multiple regression analysis was conducted. In general, all items were in between agree and strongly agree (Table 4). Table 5 shows the regression model was statistically significant, $F(2, 367) = 170.00$, $p < .001$, suggesting that the combination of predictors significantly explained variance in Exam-Oriented Practice. The model accounted for close to 48.2% of the variance in the outcome variable ($R = .694$, $R^2 = .482$, Adjusted $R^2 = .479$). Table 6 shows with a root mean square error (RMSE) of 0.472. Both Exam's Stakes ($\beta = 0.319$, SE = 0.060, t = 5.28, $p < .001$) and Exam's Format ($\beta = 0.583$, SE = 0.058, t = 10.04, $p < .001$) were statistically significant positive predictors of Exam-Oriented Practice. Standardized beta coefficients indicated that Exam's Format ($\beta = .496$) had a stronger contribution to the model than Exam's Stakes ($\beta = .261$). Table 7 shows significant main effects of both Exam's Format and Exam's Stakes on the dependent variable. The effect of The effect of Exam's Stakes was statistically significant, $F(1, 367) = 27.80$, $p < .001$, with a small-to-moderate effect size (partial $\eta^2 = .07$, $\omega^2 = .05$). The effect of Exam's Format was also statistically significant, $F(1, 367) = 100.80$, $p < .001$, with a large effect size (partial $\eta^2 = .22$, $\omega^2 = .20$). These results indicate that while both factors contributed to variance in outcomes, Exam's Format explained a substantially greater proportion of variance compared to Exam's Stakes. Diagnostic checks indicated no major violations of regression assumptions. The Durbin–Watson statistic (DW = 1.93, p = .516) in Table 8 suggested no autocorrelation in the residuals. Variance inflation factors (VIF = 1.73 for both predictors) and tolerances (0.579) confirmed that multicollinearity was not a treat in Table 9. Table 10 shows Cook's distance values were low, indicating no influential outliers affecting model estimates. The regression model significantly explained variance in the outcome variable, $R^2 = .482$, $F(2, 367) = 170$, $p < .001$, indicating that approximately 48% of the variance in Y was accounted for by the predictors and Exam's Formats serve as stronger predictor compared to Exam's Stakes in determining teacher's Exam-Oriented Practice.

Table 4: List of Item and Analysis of Mean

Item No.	Item	Mean	sd
P1	Information about the exam format is important for achieving learning objectives.	4.32	0.72
P2	The exam format should be discussed during lessons.	4.44	0.66
P3	Students prefer questions that follow the exam structure.	4.31	0.74
P4	Exam questions are used as practice during lessons.	4.29	0.69
P5	The exam format used as a basis for preparing teaching and learning materials.	4.25	0.72
Q1	Exams have a high impact on students' future.	4.27	0.73
Q2	Exams have a high impact on society's expectations.	4.35	0.69
Q3	Exams have a high impact on parents' hopes.	4.41	0.62
Q4	Exams have a high impact on students' chances of further study.	4.49	0.59
Q5	Exams have a high impact on scholarship opportunities.	4.48	0.60
Q6	Exams have a high impact on teachers' job performance.	4.23	0.71
Q7	Exams have a high impact on the school's reputation.	4.45	0.62
Q8	Exams have a high impact on improving students' family live.	4.27	0.74
R1	My teaching activities often focus on preparing students for exams.	4.05	0.82
R2	My teaching activities often focus on exam-type summative assessments.	4.00	0.78
R3	My teaching activities often use past exam papers.	3.83	0.89
R4	My teaching often focuses on topics likely to appear in exams.	3.83	0.90
R5	My teaching often focuses on drill and practice with exam questions.	3.94	0.87
R6	My teaching activities often refer to exam formats.	4.12	0.81
R7	I often conduct discussions on exam-answering techniques.	4.12	0.81
R8	I often conduct discussions about exam forecast/predicted questions.	3.73	0.97
R9	I often spend time motivating students to face exams.	4.12	0.74
R10	I often advise students to keep working hard to achieve good exam results.	4.40	0.66
R11	I often advise students to focus on exams so that they'll study.	4.28	0.74

Item No.	Item	Mean	sd
R12	My teaching activities often focus on exam-type formative assessments.	4.05	0.75

the actual survey is in Malay

Table 5: Model Fit Measures

					Overall Model Test			
Model	R	R²	Adjusted R²	RMSE	F	df1	df2	p
1	0.694	0.482	0.479	0.472	170	2	367	< .001

Table 6: Model Coefficients – MinEOP(Exam-Oriented Practice)

			95% Confidence Interval				
Predictor	Estimate	SE	Lower	Upper	t	p	Stand. Estimate
Intercept	0.129	0.2174	-0.298	0.557	0.594	0.553	
MinStakes	0.319	0.0604	0.200	0.437	5.275	< .001	0.261
MinFormat	0.583	0.0580	0.468	0.697	10.040	< .001	0.496

Table 7: Omnibus ANOVA Test and Effect Size

	Sum of Squares	df	Mean Square	F	p	Partial η^2	ω^2
MinStakes	6.26	1	6.260	27.8	< .001	0.070	0.054
MinFormat	22.67	1	22.671	100.8	< .001	0.216	0.201
Residuals	82.55	367	0.225	-	-	-	-

Table 8: Durbin–Watson Test for Autocorrelation

Autocorrelation	DW Statistic	p
0.0346	1.93	0.516

Table 9: Collinearity Statistics

	VIF	Tolerance
MinStakes	1.73	0.579
MinFormat	1.73	0.579

Table 10: Cook's Distance

			Range	
Mean	Median	SD	Min	Max
0.00296	9.92e-4	0.00741	5.37e-8	0.0687

DISCUSSION

The findings of the study showed that both predictors studied had an impact on teachers' practice of being exam-oriented. The fact that exam format was a more prominent factor than exam stakes was due to several reasons. Exams exert pressure among teachers and compelled them to frequently rely on teaching to the test as a prevalent instructional approach (Paxton, Yamazaki & Kunert, 2022). This has led to substantial shifts in classroom instruction with teachers channeling efforts towards test-focused preparation (Puspitasari & Pelawi, 2024). Consequently, high-stakes examinations have circumscribed learning predominantly to test-related materials (Owusu, 2021). Eventually, those materials align to exam's format have become the primary focus among teachers and students to facilitate their exam preparation.

Teachers also expressed that they wanted their students to be familiar with numerous text contents and types of the principle of the exams and the indication that there is interaction between teachers and learners is related to the washback of exams (Hoa, 2020). Instead of engaging in well-rounded language development, students often concentrate on mastering exam-specific tasks (Nguyen, 2023). Additionally, a heavy dependence on pre-written materials curtails creativity and individual expression in writing (Mahmood et al., 2020). This exam-driven approach often narrows the pedagogical focus of language teachers, encouraging them to prioritize short-term test outcomes over long-term communicative competence. As a result, authentic language use and learner autonomy are frequently undermined in favor of exam preparedness.

Teachers adapt their question design and instructional strategies to prepare students for new exam formats, indicating a direct influence of assessment on teaching methods (Bahrun et al., 2023). Research shows that increased practice correlates with improved student performance, suggesting that teachers may adjust their instructional practices to incorporate more practice opportunities in response to exam formats (Black et al., 2023). Familiarity with exam formats, has been shown to improve student performance, suggesting that exam structure can dictate teaching focus and preparation strategies (Neuwirt et al., 2024). Teachers often modify their curriculum to align with the requirements of specific exam formats, secure that students are adequately disposed for assessments (Thompson et al., 2023). These adaptation on exam's format highlight a strong correlation between assessment structure and instructional choices, limiting the demand for in depth understanding of meaningful learning experiences to support students' holistic development.

Washback impacts not only the content students adopt to study but also construct their overall concept of what it means to study a language (Chak, 2024). In contexts driven by high-stakes assessments, this influence can be restrictive, often narrowing students' exposure to the full range of language competencies (Desalegn, Disassa, & Kitila, 2023). The format of an exam plays a crucial role in guiding classroom instruction, for example, structured oral assessments tend to foster more active participation and clearer communication from learners (Ferine & Lestari, 2023). However, certain skills will receive less attention when resources like past exam texts are scarce (Desalegn, Disassa, & Kitila, 2023). Overall, exam-driven learning environments significantly influence both pedagogical choices and student engagement, often reinforcing narrow skill development while being shaped by social and resource-related factors.

High-stakes English tests strongly influence students' learning behaviors beyond the classroom (Desalegn et al., 2023). Teachers modify materials to align better with test demands. (Gu, 2023; Owusu, 2021). Teachers with test development experience were more aware of aligning instruction with test objectives (Athiworakun & Adunyarittigun, 2024). A well-structured preparation course enhanced confidence, but its effects diminished before the final test (Ali, 2024). Students report improved test scores due to tailored teaching materials (Gu, 2023). The curriculum, the official course documents, methods of assessment, methods of teaching and supplementary materials are innovated by the positive influences of tests. (Hoa, 2020) Published materials such as text books that similar with the exam making learning more superficial focusing on rote memorization and familiarization of exam format rather than meaningful learning (Andrews, Fullilove & Wang 2002). Such strategy promotes rote learning over deeper understanding, as a result limiting both creativity and critical thinking (Khan & Ashraf, 2023; Safa & Sheykholmoluki, 2022).

Some teachers felt pressure to modify their teaching due to the high stakes and influenced teachers' instructional choices, as they tried to prepare students for test expectations (Athiworakun & Adunyarittigun, 2024). Emphasis on memorization techniques and strategies tailored to test-taking constrains students' ability to develop comprehensive language skills (Imsa-ard, 2020). Students and parents are anxious about securing high grades. This anxiety is transferred to teachers as external pressure which leads them to focus on items that are important for public examinations (Ali & Hamid 2020). Thus shifting both teachers' and students' attention toward exam performance rather than fostering deeper engagement with language learning.

Influence from teachers, parents, peers, and society shapes students' motivation to take exams (Tsang, 2017). High social prestige associated with good exam performance led to increased parental pressure on students (Dawadi, S., 2021). Grammar, reading, and vocabulary were prioritized, while listening skills were neglected (Ali, 2024). Listening and speaking skills receive minimal attention as they are not tested in the national examination (Desalegn et al., 2023). Students who enjoy English outside of school tend to view exam

preparation as a burden (Chak, 2024). The dominance of exam-oriented teaching often sidelines communicative skills like listening and speaking, favoring the intensive instruction of skills that directly impact test scores pressured by contextual factors.

The huge test preparation industry makes publisher of materials overtly claim they are more similar to exam questions which draw attention of teachers and students to purchase them (Alderson & Hamp-Lyons, 1996). There is a large industry of guidebooks and exam preparation materials providing assurance of high scores to consumers (Ali & Hamid, 2020). Economic disparities affected test preparation, as wealthier students had access to more learning resources (CDs, tutors, newspapers) than poorer students (Dawadi, S., 2021). A test-oriented pedagogy may respond to a context-sensitive testing policy underpinned by political and economic factors such as demonstration of development and the challenges posed by the urban-rural gap (Ali & Hamid, 2020). Resource disparities among schools affected equitable implementation national exam (Puspitasari & Pelawi, 2024). High-stakes exams exacerbate educational inequalities (Caner, 2020; Khan & Ashraf, 2023). The commercialization of exam preparation not only distorts teaching priorities but also deepens social inequities, as affluent students gain advantages unavailable to their less privileged peers.

Assessment fairness is a concern, as School-Based Assessment (SBA) is conducted by individual schools, leading to inconsistencies (Auladi, 2024). Additionally, students' conceptualizations of fairness in assessments, reveal that they associate fairness with equal educational opportunities, which may not always be present in SBA (McCracken et al., 2024). Students lacked access to modern teaching tools, relying solely on textbooks for preparation (Ali, 2024). Limited professional development on SBA is also a major challenge (Auladi, 2024). Yet, traditional exams are often viewed as more straightforward and less time consuming compared to the multifaceted nature of SBA, which requires ongoing assessment and feedback (Alkan et al., 2024). The shortcomings of SBA and traditional exams underscore the urgency for holistic and sustainable language assessment approaches that value continuous learning and equitable access over rote performance metrics.

The study found a lack of alignment between test items and the prescribed textbook (Ahmed & Sikder, 2023). Curriculum alignment with exams remains weak despite policy claims (Caner, 2020). Furthermore, lack of teacher training on assessment literacy limited teachers' ability to implement intended washback (Athiworakun & Adunyarittigun, 2024). Continuous assessment of listening and speaking skills should be incorporated to improve test validity (Ahmed & Sikder, 2023). Assessment literacy would help teachers integrate test preparation without sacrificing communicative teaching goals (Athiworakun & Adunyarittigun, 2024). Research indicates that closed-ended formats often yield higher scores compared to open-ended formats can lead to an emphasis on lower-order thinking skills (Breuer et al., 2023).

Language test's reliability could be enhanced by using formative assessment (Ahmed & Sikder, 2023). SBA is assented as better align with language curriculum and capable in nurturing higher-order thinking skills (Auladi, 2024). Further variables that influence the practice of language learning or test preparation should be put into consideration by future washback researchers (Chou, 2018). Methods using questionnaires, observation, or analysis of textbooks could be carried out to investigate more about washback effects (Chou, 2017). Also, the future educational assessment is composed for consequential transformation, specifically in contrast to conventional high-stakes testing practice.

Considering innovation with ethical deliberation will be essential in designing the future of educational assessment (Ho, 2024). A more comprehensive evaluation of student learning outcomes could be achieved by integrating traditional and innovative method (Weng et al., 2024). Artificial Intelligence (AI) tools enhance the accuracy and effectiveness of assessment rubrics and aligning it with curricular outcomes and adding inclusive learning experiences (Sánchez et al., 2024). The merging of AI into educational assessments drives for innovative designs that emphasis on critical thinking and humanizing input and diminishing academic misconduct (Furze et al., 2024). The potential for biased outcomes and academic integrity remains while the move towards technology-enhanced educational assessment supply numerous advantages.

CONCLUSION

The study concludes that both exam format and stakes significantly shape teachers' assessment decisions, with exam format employing the strongest influence, explaining nearly half of the variance in instructional decisions. This provides a quantitative proof using regression analysis that both predictors directly influence teachers' pedagogical approaches. It also offers insights for stakeholders especially policymakers and curriculum developers by highlighting how assessment format should be design that drives teaching behavior and balancing the needs of holistic and unbiased assessment systems. This research underlines the significant washback effect of exam's format and stakes on teaching, providing to assessment literacy and policy development, while also highlighting to future demand for AI integration, comprehensive and authentic language assessment systems that better apprehend real communicative competence and diminished negative washback

ETHICAL CONSIDERATIONS

The questionnaire design and adminstration followed ethical guidelines, preserving respondents anonymity, privacy and voluntary engagement.

CONFLICT OF INTEREST

No conflicts of interest have been identified by the authors in relation to the research, authorship, or publication of this work.

FUNDING

REFERENCES

1. Abate, T., & Mishore, E. (2024). Alignment Analysis between Teacher-made Tests with the Learning Objectives in a Selected School of Central Regional State of Ethiopia. Heliyon. https://doi.org/10.1016/j.heliyon.2024.e31869

2. Aftab, A., Qureshi, S. & William, I. (2014). Investigating the Washback Effect of the Pakistani Intermediate English Examination. International Journal of English & Literature. Vol.5(7). 149-154 DOI: 10.5897/IJEL2013.0521

3. Aidil, M., Jabar, J., Sufian, A., Fauziana, N., & Abdul, F. (2023). Employing Fuzzy Delphi Techniques to Validate the Components and Contents of E-Learning Antecedents and Usage Behavior Towards E-Learning Performance. European Journal of Educational Research, 12(1), 467–480. https://doi.org/10.12973/eu-jer.12.1.467

4. Ahmed, I., & Sikder, L. B. (2023). Measuring the validity of a high-stakes English test in Bangladesh: An empirical approach. International Journal of Research and Innovation in Social Science. 7(5), 1349-1369. https://doi.org/10.47772/IJRISS.2023.70603

5. Alderson, J. C. & Hamp-Lyons, L. (1996). TOEFL Preparation Courses: A Study of Washback. Language Testing. 13:280. DOI:10.1177/026553229601300304

6. Ali, M. M.& M. Obaidul Hamid, O. (2020): Teaching English to the Test: Why Does Negative Washback Exist within Secondary Education in Bangladesh?, Language Assessment Quarterly, DOI: 10.1080/15434303.2020.1717495

7. Ali, R. E. (2024). Washback in English Education: Insights from Libyan Ninth Graders. Journal of English Studies in Arabia Felix, 3(2), 35-43. DOI: 10.56540/jesaf.v3i2.94

8. Alkan, Ö. K., Aksoy, N., Kulaksiz, T., Kaplan, H. A., Durmaz, B. N., Özcan, M., & Kalkavan, B. (2024). A multi-feedback system integrated simulation-based teacher training to scaffold pre-service teachers' teaching skills: A phenomenological approach. Education and Information Technologies. https://doi.org/10.1007/s10639-024-12657-4

9. Alla Baksh, M. A., Mohd Sallehhudin, A. A., & Siti Hamin, S. (2019). Examining the Factors Mediating the Intended Washback of the English Language school-based Assessment: Pre-service ESL Teachers' Accounts. Pertanika J. Soc. Sci. & Hum. 27 (1): 51 – 68

10. Allen, D. (2016). Japanese Cram Schools and Entrance Exam Washback. The Asian Journal of Applied Linguistics. Vol. 3(1).54-67

11. Allen, D., Yamamura, K., Meguro, S., & Nakamura, T. (2024). Washback from Four-Skills English Exams in the Japanese Secondary School Context. Language Assessment Quarterly, 1–33. https://doi.org/10.1080/15434303.2024.2440895

12. Andrews, S., Fullilove, J. & Wong, Y. (2002). Targeting Washback: A Case Study. System. Vol.30. 207-223

13. Athiworakun, C., & Adunyarittigun, D. (2024). Teacher factors as a mediator between teaching and testing at the higher education level. rEFLections, 31(3), 944-965.

14. Auladi, I. R. (2024). The washback of the abolition of the national examination in Indonesia. Indonesian Journal of Education and Pedagogy, 1(1), 48-63. https://doi.org/10.61251/ijoep.v1i1.50

15. Baharuddin, K. A., Abdul Rahim, A. F., Abdul Aziz, Y. F., Abdul Manap, R., Omar, A. M., Singmamae, N., et al. (2020). Recommendations for high-stakes examinations for public Malaysian medical faculties during the COVID-19 pandemic. Education in Medicine Journal, 12(4), 25–31. https://doi.org/10.21315/eimj2020.12.4.4

16. Bahrun, B., Maulana, R. I., Muslem, A., & Yulianti, Y. (2023). Assessment design, learning strategies and obstacles in facing Computer-Based Madrasah Exam on the English subject. Studies in English Language and Education, 10(2), 884–906. https://doi.org/10.24815/siele.v10i2.31954

17. Black, W. K., Matz, R. L., Mills, M., & Evrardsj, A. (2023). Quantifying grade benefits of study: Practice makes better. Physical Review, 19(1). https://doi.org/10.1103/physrevphyseducres.19.010139

Open Access

18. Bond, T. G., & Fox, C. M. (2015). Applying the Rasch model: Fundamental measurement in the human sciences (3rd ed.). Routledge.

19. Bordel, B., Alcarria, R., & Robles, T. (2023). Does the exam format influence the students' academic results? multiple-choice exams vs short-answer type exams in engineering courses. ICERI Proceedings. https://doi.org/10.21125/iceri.2023.1036

20. Breuer, S., Scherndl, T., & Ortner, T. M. (2023). Effects of response format on achievement and aptitude assessment results: multi-level random effects meta-analyses. Royal Society Open Science, 10(5). https://doi.org/10.1098/rsos.220456

21. Chak, M. C. (2024). Washback in Language Learning Strategies Under High-Stakes Language Testing: A Study of the Hong Kong Secondary System. rEFLections, 31(1), 1-24

22. Chan, K. L. R. 2020. Washback in Education: A Critical Review and Its Implications for Language Teachers. Journal of Foreign Language Education and Technology, Vol.5(1). 109-124.

23. Cheng, L., Sun, Y. & Ma, J. (2015). Review of washback research literature within Kane's argument-based validation framework. Lang. Teach. Vol.48(4), 436–470. doi:10.1017/S0261444815000233

24. Cheng, L-Y. (1997). How Does Washback Influence Teaching: Implication for Hong Kong. Language Education. Vol.11(1). 38-54

25. Chou, M-H. (2017). Impacts of the Test of English Listening Comprehension (TELC) on teachers and teaching in Taiwan. Asian-Pacific Journal of Second and Foreign Language Education. Vol.2(5) DOI 10.1186/s40862-017-0028-9

26. Chou, M-H. (2019): The Impact of The English Listening Test In The High-Stakes National Entrance Examination On Junior High School Students And Teachers, International Journal of Listening, DOI: 10.1080/10904018.2019.1659142

27. Christensen, K. B., Makransky, G. & Horton M. (2017). Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations. Appl Psychol Meas. 41(3):178-194. doi: 10.1177/0146621616677520. Epub 2016 Nov 16. PMID: 29881087; PMCID: PMC5978551.

28. Clark, T., & Yu, G. (2022). Test preparation pedagogy for international study: Relating teacher cognition, instructional models and academic writing skills. Language Teaching Research, 29(2), 726-750. https://doi.org/10.1177/13621688211072381 (Original work published 2025)

29. Dawadi, S. (2021). Factors affecting washback of a high-stakes English as a foreign language test. TESL-EJ: The Electronic Journal for English as a Second Language, 25(3). https://oro.open.ac.uk/79796/

30. DeCarlo, L. T. (2024). Fused SDT/IRT Models for Mixed-Format Exams. Educational and Psychological Measurement. https://doi.org/10.1177/00131644241235333

31. Desalegn, G., Disassa, R., & Kitila, T. (2023). The influence of high-stakes English examinations on students' out-of-classroom English learning practices: A comparative study. Education Research International, Article 1108951, 1–15. https://doi.org/10.1155/2023/1108951

32. Ferine, M., Nauvalina, R., & Lestari, D. W. D. (2023). Impact of Structured Oral Examination 0n Students' Learning: A Qualitative Research In Faculty of Medicine Universitas Jenderal Soedirman. Medical and Health Journal, 2(2), 39. https://doi.org/10.20884/1.mhj.2023.2.2.834

33. Fernández Sánchez, A., Lorenzo Castiñeiras, J. J., & Sánchez-Bello, A. (2024). Navigating the Future of Pedagogy: The Integration of AI Tools in Developing Educational Assessment Rubrics. European Journal of Education. https://doi.org/10.1111/ejed.12826

34. Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. Australasian Journal of Educational Technology. https://doi.org/10.14742/ajet.9434

35. Gaber, I., Kirsh, A., & Statter, D. (2023). Studied Questions in Data Structures and Algorithms Assessments. Annual Conference on Innovation and Technology in Computer Science Education. https://doi.org/10.1145/3587102.3588843

36. Goble, R. (2024). Language Learners' Transferable Skills and Narrative Competence in U.S. Career Advising Appointments: Multilingual Development Beyond Higher Education. Journal of Language Identity and Education, 1–19. https://doi.org/10.1080/15348458.2024.2375280

37. Greer, T., Mills, M. T., & Lapka, S. (2024). The influence of linguistic profiling on perceived employability: a new application of the systems theory framework of career development. Career Development International. https://doi.org/10.1108/cdi-07-2023-0231

38. Grover, P., & Pandya, B. (2024). An Analytical Study of Acquiring English Communication Skills for Employability being Affected by Societal Factors and Own Personality Traits of Learners of Technical Graduates in India. Evolutionary Studies in Imaginative Culture, 1478–1484. https://doi.org/10.70082/esiculture.vi.1544

39. Ho, A. D. (2024). Artificial Intelligence and Educational Measurement: Opportunities and Threats. Journal of Educational and Behavioral Statistics. https://doi.org/10.3102/10769986241248771

40. Linacre, J. M. (2005). A user's guide to Winsteps: Rasch-model computer programs. Winsteps.com

41. Manakandan, S. K., Rosnah, I., Mohd Ridhuan, J., & Priya, R. (2017). Pesticide applicators questionnaire content validation: A fuzzy delphi method. *The Medical Journal of Malaysia*, 72(4), 228–235. https://pubmed.ncbi.nlm.nih.gov/28889134/

42. McCracken, M., Bostic, J. D., & Folger, T. (2024). Middle School Students' Conceptualizations and Reasoning about the Fairness of Math Tests. Techtrends. https://doi.org/10.1007/s11528-024-00988-5

43. Mohd Nazri Latiff, A. & Rosnah, M. S. (2018). The Washback Effect of School-Based Oral Evaluation Performance and Its Causal Relationship with Fear of Negative Evaluation. *Language Literacy*, Vol.2 (1) e-ISSN: 2580-9962 | p-ISSN: 2580-8672 DOI: 10.30743/

44. Neuwirt, H., Eder, I. E., Gauckler, P., Horvath, L., Koeck, S., Noflatscher, M., Schaefer, B., Simeon, A., Petzer, V., Prodinger, W. M., & Berendonk, C. (2024). Impact of familiarity with the format of the exam on performance in the OSCE of undergraduate medical students – an interventional study. *BMC Medical Education*, 24. https://doi.org/10.1186/s12909-024-05091-0

45. Ramalingam, M., Halili, S. H., & Siraj, S. (2019). Experts' Agreement of the Personalized m-Learning Curriculum Model Based on Fuzzy Delphi Method. 407–420. https://doi.org/10.22437/IRJE.V3I2.7915

46. Tahir, M. E., Hamdan, H., Gaddal, A., & Gasmalla, H. E. E. (2023). Fostering formative assessment: teachers' perception, practice and challenges of implementation in four Sudanese medical schools, a mixed-method study. *BMC Medical Education*, 23(1). https://doi.org/10.1186/s12909-023-04214-3

47. Thompson, A. R., Lowrie, D. J., & Ubani, M. (2023). The Effect of Histology Examination Format on Medical Student Preparation and Performance: Stand-Alone Versus Integrated Examinations. *Medical Science Educator*, 33(1), 165–172. https://doi.org/10.1007/s40670-023-01731-0

48. Tsang, C. L. H. (2017). Examining Washback on Learning from a Sociocultural Perspective: The Case of a Graded Approach to English Language Testing in Hong Kong [Master Thesis]. University College London: British Council Hong Kong.

49. Weng, X., Xia, Q., Gu, M., Rajaram, K., & Chiu, T. K. F. (2024). Assessment and learning outcomes for generative AI in higher education: A scoping review on current research status and trends. *Australasian Journal of Educational Technology*. https://doi.org/10.14742/ajet.9540

50. Ye, X., Yu, Y., Qie, Y., & Hu, Z. (2024). The impact of 24-h take-home exam on language learning and teaching on the China campus of a British university. *Language Testing in Asia*, 14(1). https://doi.org/10.1186/s40468-024-00295-z

51. Zakaria, A., & Aryadoust, V. (2024). From Colonialism to Self-determination: A Historical Study of Language Testing and Institutional Policy. *Applied Linguistics*. https://doi.org/10.1093/applin/amae057

52. Wang, J., Zheng, Y., & Zou, Y. (2024). Face validity and washback effects of the shortened PTE Academic: insights from teachers in Mainland China. *Language Testing in Asia*, 14(1). https://doi.org/10.1186/s40468-024-00302-3

53. Zarei, M. A., Alibakhshi, G., & Nezakatgoo, B. (2024). Strategies employed by EFL teachers to cope with language learners' classroom anxiety. *TESOL Journal*. https://doi.org/10.1002/tesj.843

54. Zhou, Y., Wieringa, T. H., Brouwer, J., Diemers, A. D., & Bos, N. A. (2023). Challenges to acquire similar learning outcomes across four parallel thematic learning communities in a medical undergraduate curriculum. *BMC Medical Education*, 23(1). https://doi.org/10.1186/s12909-023-04341-x