LLT Journal, e-ISSN 2579-9533, p-ISSN 1410-7201, Vol. 28, No. 2, October 2025, pp. 882-901

LLT Journal: A Journal on Language and Language Teaching
http://e-journal.usd.ac.id/index.php/LLT
Sanata Dharma University, Yogyakarta, Indonesia

EFFECTS OF IDIOMS ON TEACHER RATINGS OF L2 WRITING QUALITY AND LEXICAL DIVERSITY, AND ASSESSMENT CHALLENGES

Eihab Abu-Rabiah
Kaye Academic College of Education, Israel
Ben-Gurion University of the Negev, Israel
aehab@kaye.ac.il
correspondence: aehab@kaye.ac.il
https://doi.org/10.24071/llt.v28i2.12112
received 13 April 2025; accepted 11 September 2025

Abstract
Idioms are often seen as signs of advanced language proficiency, yet their role in writing assessment is given little consideration. Most research has focused on how idioms are processed cognitively, leaving open questions about how they affect teacher judgments, particularly in second language writing. This study explores how idioms affect teacher ratings of lexical diversity and writing quality, while also highlighting the challenges teachers face in evaluating them. Using a mixed-methods design, twenty Hebrew teachers assessed fifteen essays written by native Arabic speakers, along with revised versions of each essay that included a frequent idiom (30 essays total). The essays spanned CEFR levels A2 to B2 and were randomly ordered. Teachers, whether native speakers of Arabic or Hebrew, rated both lexical diversity and writing quality of each essay on a 10-point scale. Six teachers also participated in semi-structured interviews, which were subjected to thematic analysis. The findings indicate that essays with idioms were rated significantly higher in both writing quality and lexical diversity, regardless of the rater's first language. The interviews highlighted several recurring challenges in assessing idioms, including contextual appropriateness, idiom transparency, frequency of use, grammatical accuracy, and variation between standard and colloquial forms.

Keywords: formulaic language, Hebrew as a second language, L2 writing quality assessment, lexical diversity, teacher rating

Introduction
Idioms, which are widely considered to be a sign of high language proficiency, improve lexical knowledge, fluency, and naturalness in both written and spoken communication (Fernando, 1996; Mehdi, 2024; Vasiljevic, 2015). However, despite their recognized importance (Paquot, 2018), research on idioms has focused more on second language acquisition and processing rather than on assessment. In particular, their role in second language (L2) writing assessment remains underexplored. Most studies have concentrated on the cognitive processing of idioms (Cacciari & Tabossi, 2014; Milburn et al., 2021), their link to global

This work is licensed under CC BY-SA.
Creative Commons Attribution-ShareAlike 4.0 International License

proficiency (Al-Kadi, 2015; Ellis et al., 2008), or differences in idiom use and comprehension between native and non-native speakers (Kashiha & Chan, 2015; Senaldi & Titone, 2024). This leaves an important gap: we still know little about how the inclusion of idioms in L2 writing influences teachers' judgments of writing quality and lexical diversity.

Idioms, fixed or semi-fixed multi-word expressions with meanings that cannot be inferred from their parts (Gramley et al., 2020; Wagner, 2020), are markers of phraseological competence, an under-assessed dimension of L2 writing (Paquot, 2018). They enrich writing by adding depth, emotional resonance, and cultural authenticity (Fernando, 1996; McGlone et al., 1994; Mehdi, 2024). Yet, their semantic opacity and cultural specificity make them difficult for L2 learners to master and for teachers to evaluate. Limited exposure, figurative language challenges, and first language (L1) transfer all increase the likelihood of inappropriate use (Hajiyeva, 2024; Listyani & Thren, 2023; Siyanova-Chanturia et al., 2011; Ta'amneh, 2021). This dual nature—idioms as both assets and risks—means that raters may reward, ignore, or penalize idiomatic usage inconsistently. Furthermore, their fixedness may artificially reduce variation in automated measures of lexical diversity (Crossley & McNamara, 2012), underscoring the need to examine how teachers, particularly in multilingual settings, judge idioms in L2 writing.

To address this gap, the present study investigates the effects of idioms on teachers' evaluations of L2 Hebrew writing quality and lexical diversity. Unlike previous research that has primarily examined idioms from the perspective of learners' cognitive processing or comprehension, this study situates idioms in an authentic assessment context, where teachers evaluate actual student writing. It also explores the specific difficulties teachers face when judging idiomatic usage. In doing so, the study considers whether teachers' linguistic backgrounds (native Hebrew vs. native Arabic) (Zhang & Elder, 2011) shape their evaluations and contributes to the broader call to recognize phraseological competence as a core component of L2 writing proficiency (Paquot, 2018). Specifically, it addresses two questions:

1. How do Hebrew teachers who are native speakers of Arabic or Hebrew assess lexical diversity and writing quality when idioms are included?
2. What challenges do teachers encounter when evaluating idiomatic language in L2 writing?

Literature Review

Idiom definitions

Idioms are notoriously difficult to define, and scholars continue to debate their essential characteristics (Espinal & Mateu, 2019; Rafatbakhsh & Ahmadi, 2019). At the heart of this debate lies the principle of non-compositionality, the idea that an idiom's meaning cannot be predicted from its individual parts. While some definitions emphasize their fixedness and conventionalization (Casas & Campoy, 1995; Wagner, 2020), others highlight their semantic unity, describing idioms as "a complex lexical item… shorter than a sentence and with a meaning that cannot be derived" (Gramley et al., 2020, p. 44). These definitional ambiguities are not merely theoretical but they also have direct pedagogical and assessment implications. If teachers themselves cannot agree on what counts as an idiom, assessment risks

inconsistency, with idiomatic expressions treated as memorized chunks by some and as markers of sophistication by others. This uncertainty underscores the need for empirical research into how idioms are operationalized in actual L2 writing assessment contexts.

Idioms in writing

Idioms play a significant role in writing, where they enhance vocabulary, pragmatic awareness, and style (Bestgen & Granger, 2014; Saberian & Fotovatnia, 2011; Vasiljevic, 2015). Their use extends beyond casual conversation and can appear even in formal genres such as academic writing (Miller, 2020). Theoretically, Sinclair's (1991) *idiom principle* and Lewis's (1993) *lexical approach* both emphasize the importance of formulaic language in producing natural and fluent texts. Empirically, studies show that writers who skillfully use idioms can achieve greater expressiveness and rhetorical impact (Fernando, 1996; McGlone et al., 1994). Yet, idioms can also complicate assessment. As they are fixed expressions, they may reduce lexical diversity or the number of different words used in a text, as opposed to lexical repetition (Abu-Rabiah, 2020, 2023; Malvern et al., 2004; McCarthy & Jarvis, 2007) in automated measures (Crossley, 2020; Crossley & McNamara, 2012), raising questions about whether human judgments better capture idioms' contribution to lexical diversity. This tension between idioms' stylistic value and their potentially distorting effects on lexical diversity assessment makes it crucial to study how teachers themselves weigh idioms when evaluating L2 writing.

Processing and application of idioms by both native and non-native speakers

Idioms are processed and used differently by native and non-native speakers due to differences in familiarity, contextual awareness, and language proficiency (Al-Kadi, 2015; Cacciari & Tabossi, 2014). Native speakers typically enjoy idiom superiority, processing idioms more quickly and flexibly than literal equivalents (Carrol & Conklin, 2020). By contrast, L2 learners often struggle with semantic opacity, figurative meanings, and L1 transfer, which lead to slow recognition, literal translations, and inappropriate usage (Anjarini & Hatmanto, 2022; Hajiyeva, 2024; Listyani & Thren, 2023; Siyanova-Chanturia et al., 2011; Suhodolli & Lama, 2024). These difficulties are not only cognitive but also cultural, requiring sociolinguistic awareness to deploy idioms appropriately (Hajiyeva, 2024; Ta'amneh, 2021). Crucially, such challenges spill over into assessment. When idioms are misused, whether due to semantic misinterpretation, contextual mismatch, or register violations, teachers must decide whether to penalize or overlook them. The inconsistency of these judgments illustrates why idioms remain underexplored yet highly consequential in L2 writing evaluation.

Despite the well-established connection between idioms and language proficiency, little is known about how they affect teachers' assessments of L2 writing. The importance of idioms in real-world assessment contexts has been overlooked in favor of focusing on the cognitive processing of idioms and the differences in idiom usage and comprehension between native and non-native speakers. Furthermore, neither formulaic language in general nor idioms in particular have been directly addressed in any of the many studies on writing in Hebrew as an L2 (Abu-Gweder, 2023; Abu-Rabiah, 2025a; Abu-Rabiah, 2025b;

Abu-Rabiah, 2025c; Abu-Rabiah et al., 2023; Haskel-Shaham et al., 2018; Henkin, 2020). Thus, a mixed-methods methodology is used in this study to examine the ways in which idioms affect teachers' assessments of lexical diversity and writing quality. Through the integration of semi-structured interviews and quantitative data from essay ratings, this study attempts to fill this knowledge vacuum and investigate the real-world difficulties teachers have when evaluating idiomatic language.

Method

Research design

Drawing on a mixed-methods design (Johnson et al., 2007; Plano Clark, 2017), this study explored how weaving idiomatic expressions into L2 writing influenced teachers' evaluations of both writing quality and lexical diversity. The study design also identified difficulties in evaluating idioms in L2 writing by combining quantitative analysis (Likert scale ratings) with qualitative insights (semi-structured interviews). A mixed-methods approach was especially appropriate since it made it possible to record both quantifiable results and the underlying causes of them. While qualitative interviews supplied insights into the interpretative difficulties and contextual aspects that numbers alone could not show, quantitative evaluations on Likert scales demonstrated whether idioms consistently impacted teacher perceptions.

Participants: Essay evaluators and interview participants

Twenty Hebrew language teachers participated in rating the essays. The researcher's academic network was used to recruit them. Prior to participation, every participant provided their informed consent. The researcher described the aim of the study, the steps involved, and the voluntary nature of involvement before starting the rating task. The teachers were told that their answers would remain confidential, that they could cease participating at any time without facing any consequences, and that the information would only be used for research. Table 1 below provides a summary of the evaluators' demographic characteristics.

Table 1. Demographic characteristics of essay evaluators

Characteristic	N	Percentage
Gender	20	100
Female	10	50
Male	10	50
Native language	20	100
Hebrew	11	55
Arabic	9	45
Teaching level	20	100
College/university	15	75
High school	3	15
Primary school	1	5
Middle school	1	5
Age		
Mean (SD)		44.9 (11.4)
Range		26–65

Characteristic	N	Percentage
Teaching experience		
Mean (SD)		18.2 (12.6)
Range		2–46

Six teachers in all took part in the semi-structured interviews. The teachers from each of the two L1 groups were contacted after they were chosen at random. The first three respondents from each group were selected to take part. Table 2 below provides a summary of the participants' demographic information.

Table 2. Demographic characteristics of interview participants

Characteristic	N	Percentage	Detail	Value
Gender	6	100	Male	3
			Female	3
Native language	6	100	Arabic	3
			Hebrew	3
Teaching level	6	100	College/university	5
			High school	0
			Primary school	1
Age			Mean (SD)	41.5 (11.2)
			Range	28.0–57.0
Teaching experience			Mean (SD)	15.7 (8.5)
			Range	5–26

Corpus

The essays for this study were drawn from a larger collection of 156 argumentative essays. These were all written as part of an academic college admission exam in Israel designed to measure Hebrew proficiency of the students. The essay prompt was "Should parents compensate children for assisting with household chores, such as cleaning and caring for younger siblings?" To ensure a fair representation of different proficiency levels, fifteen essays from the full corpus were selected, with a total length of 1,686 words ($M \approx 112$ words per essay). These were specifically chosen to reflect the three proficiency levels identified by the Common European Framework of Reference for Languages (CEFR): pre-intermediate (A2), intermediate (B1), and upper-intermediate (B2). To account for variability and focus on the most representative work for each level, the five essays with the highest scores from each proficiency category were selected.

Idiom selection

To include idioms suitable for intermediate-level Hebrew learners, 75 idioms were initially selected from an extensive Hebrew idiom book. Idioms were defined following Cacciari as "strings of words whose figurative meaning does not necessarily derive from that of the constituent parts" (2014, p. 267). There are no documented frequency lists for Hebrew idioms, so a Google search function was used to quantitatively refine the collection. The search term was ["exact idiom phrase" site:il]. Idioms that yielded more than 50,000 search results were deemed

sufficiently common. From this condensed list, the 15 idioms with the highest search results were selected to guarantee relevance and familiarity within the target language community. These idioms are shown in Table 3.

Table 3. Fifteen Hebrew idioms selected for the study

Hebrew idiom	Transliteration	Literal meaning	Idiomatic meaning
למראית עין	Lemar'it ayin	To the appearance of the eye	Superficially / seemingly / on the surface
מוטל בספק	Mutal be-safek	Placed in doubt	Doubtful/uncertain
בכובד ראש	Be-koved rosh	With a heavy head	Seriously / with gravity
פרשת דרכים	Parashat drakhim	A road junction	A turning point / critical decision moment
בעת ובעונה אחת	Be-et u-ve-'onah achat	At a time and at one season	Simultaneously / at the same time
לפנים משורת הדין	Lifnim mi-shurat ha-din	Beyond the letter of the law	Acting with compassion or leniency beyond strict rules
כברת דרך	Kivrat derekh	A segment of road	A considerable amount of progress
בפה מלא	Be-feh male	With a full mouth	Unreservedly/unequivocally
קורע לב	Korea's lev	Heart-tearing	Heartbreaking
אסיר תודה	Asir todah	Prisoner of thanks	Deeply grateful
קל וחומר	Kal va-chomer	Light and heavy	All the more so / even more so (a fortiori)
בד בבד	Bad be-vad	Thread with thread	Simultaneously/in parallel
גולת הכותרת	Gulat ha-koteret	The top of the headline	The highlight/the crown jewel
בראש ובראשונה	Be-rosh u-be-rishonah	First and foremost	Primarily / above all
מאחורי הקלעים	Me'achorei ha-klayim	Behind the curtains	Behind the scenes

Essay preparation

For each of the fifteen selected essays, two versions were prepared: an original version without any idioms and a modified version that started with an idiom. In most cases, adding the idiom didn't disrupt the essay's structure. However, minor syntactic adjustments were occasionally made to the first sentence to ensure it flowed smoothly. Before they were evaluated, all essays were proofread and corrected for spelling mistakes. This step was taken to prevent any confounding variables. It's a well-documented phenomenon in assessment research that seemingly minor details, like spelling accuracy, can create a halo effect, influencing how raters judge other aspects of a writer's language skills (O'Grady, 2023). By addressing these issues in advance, the systematic editing procedure made it easier to pinpoint and quantify the precise effect of using an idiom on the final writing assessment.

Evaluation procedure and randomization

Based on an online questionnaire, the evaluation procedure asked participants to score each essay on a Likert scale of 1 to 10 for writing quality and lexical

diversity (Joshi et al., 2015). For writing quality, no rigid rubric was enforced; raters relied on their own pedagogical judgment, as they would in everyday classroom settings. Lexical diversity, however, came with a clear definition to ensure that everyone was reading from the same playbook. Each essay was reviewed within a strict two-minute window, making the entire evaluation process clock in at just about an hour. The premise that accurate assessments of lexical diversity can be made quickly with little information and direction justifies the evaluation procedure, in which raters were given two minutes per essay (Meara, 2014). Notably, formal training is not a prerequisite for reliably evaluating lexical diversity; raters tend to show a high level of consistency nonetheless (Jarvis, 2017). As for the two-minute cap, it aligns well with the everyday realities of L2 instruction, where assessing writing, often quickly and intuitively, is a routine part of the job. Alongside essay ratings, basic demographic data, such as participants' age, gender, type of institution, teaching experience, and native language, were also collected through the questionnaire.

Pre-intermediate (A2), intermediate (B1), and upper-intermediate (B2) proficiency levels, as well as the presence or lack of idioms, were balanced by presenting the essays in a randomized order to reduce bias. By distributing the essays in a random order, as illustrated in the scatter plot (Figure 1), the risk of systematic bias was decreased, making the evaluation process more unbiased and trustworthy and guaranteeing that all variables (idiomatic usage and proficiency level) were equally represented throughout the evaluation sessions.

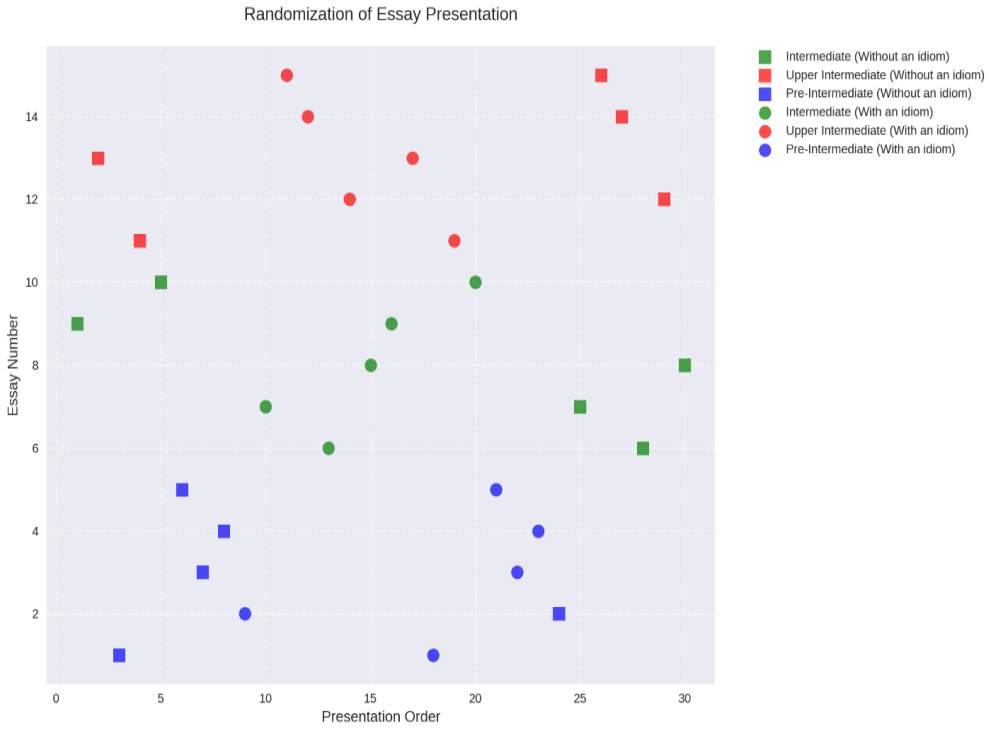


Figure 1. Randomized presentation order of essays across proficiency levels and idiomatic versions

Semi-structured interviews

Six participating teachers participated in semi-structured interviews (Adams, 2015) to augment the quantitative results. Once the teachers had finished the online

survey, the researcher conducted the interviews in person or over the phone. This order made sure that their answers to the questionnaire were unaffected by the interviews. Two main study topics were the focus of each 30-minute interview: (1) how idioms impacted lexical diversity and overall writing quality in L2 writing, and (2) the difficulties the teachers had when assessing idioms in L2 writing. While the second question examined evaluation challenges using qualitative methods, the first question sought to contextualize the quantitative findings. Since the teachers were not aware that the rating task had evaluated idiom impacts on evaluation, they were asked about their general perceptions of idioms and the difficulties in evaluating them, regardless of the particular essays they had graded.

The interview transcripts were analyzed using Braun and Clarke's (2006) six-phase thematic analysis (see also Kiger & Varpio, 2020). Initial codes were generated inductively, tagging any segment where the teachers articulated criteria or difficulties in judging idioms (e.g., "check whether the use is in the correct context," Teacher 1; frequent/common idioms at higher levels signaling limited variety; idioms accompanied by grammatical/syntactic errors suggesting memorization rather than mastery; tolerance of nonstandard forms if native-like). Codes were then clustered by (a) conceptual focus (what aspect of evaluation they addressed, such as context, complexity, frequency/level, form/accuracy, normativity), (b) evaluative function (how the issue changed scores), and (c) recurrence across participants. This process produced the following overarching themes: *Contextual Appropriateness* (codes on register/genre fit and correct figurative meaning), *Idiom Complexity* (Transparency) (greater difficulty and lower consistency for opaque items), *Idiom Frequency* (common idioms acceptable at lower levels but signaling reduced lexical diversity at higher levels), *Grammatical/Syntactic Accuracy* (idioms accompanied by errors were seen as memorized rather than mastered, reducing authenticity and lowering scores, and *Normative vs. Colloquial Variation* (penalizing deviation from textbook forms vs. accepting widespread native-like variants). Ambiguous or overlapping codes were resolved by returning to the full interview context and privileging the rationale the teachers explicitly linked to scoring decisions.

Quantitative data from the essay ratings addressed the first research question by measuring whether idiom inclusion and rater's L1 systematically influenced perceptions of writing quality and lexical diversity. The qualitative data then contextualized these patterns, explaining why raters rewarded idioms but also struggled with consistency, as reflected in the themes identified above.

Findings and Discussion

Findings

Effects of idiom inclusion and rater L1 on L2 writing quality judgments

The effects of native language (between-subjects factor: Arabic vs. Hebrew) and idiom inclusion (within-subjects factor: with idioms vs. without idioms) on writing quality were investigated using a mixed-design ANOVA. Essays containing idioms ($M = 5.70$, $SD = 2.04$) were judged as having greater writing quality than essays without idioms ($M = 5.32$, $SD = 1.94$), according to the analysis, which showed a significant main impact of idiom inclusion ($F(1, 18) = 5.76$, $p = .027$, $\eta^2 = .24$). Idiom inclusion appears to be responsible for 24% of the variance in writing quality scores, according to the large effect size ($\eta^2 = .24$). Interview data revealed

that the teachers consistently perceived the use of idioms as a marker of advanced proficiency because it improved writing fluency, necessitated a broader understanding of semantics, and showed a greater command of the language. These quantitative results were complemented by interview findings that clarified why idioms enhanced ratings of writing quality. Teacher 1 said, "The proper use of idioms shows an advanced language… It shows high language ability." Idioms were seen to enhance expressive style and fluency, with Teacher 2 describing idiomatic lines as "sparkling" compared to more literal expressions.

Native language had little effect on writing quality, as seen by the lack of a significant main effect ($F(1, 18) = 0.29$, $p = .598$, $\eta^2 = .016$). Additionally, only 2.1% of the variance was explained by the interaction between idiom inclusion and native language, $F(1, 18) = 0.38$, $p = .545$, $\eta^2 = .021$, indicating a minor combined influence of both factors.

Effects of idiom inclusion and rater L1 on L2 lexical diversity judgments

In addition to writing quality, the study examined how idiom use and rater L1 influenced judgments of lexical diversity. The effects of native language and idiom inclusion on lexical diversity were also assessed using a mixed-design ANOVA. The findings showed that idiom inclusion had a significant main effect ($F(1, 18) = 5.27$, $p = .034$, $\eta^2 = .23$), with essays with idioms ($M = 5.87$, $SD = 2.09$) showing more lexical diversity than those without ($M = 5.35$, $SD = 2.00$). Idiom inclusion accounts for 23% of the variance in lexical diversity, as indicated by the effect size ($\eta^2 = .23$). The teachers generally agreed that idioms can signal lexical diversity, though not necessarily lexical richness. Qualitative data supported these results, with the teachers noting that idioms often expanded the perceived lexical range while also cautioning against overuse or superficial deployment. Teacher 3 pointed out, "Incorporating idioms indicates diversity, but it does not always indicate lexical richness." Idioms are typically absent from limited vocabularies, so their use suggests a broader lexical range. However, frequent use without full understanding may limit their contribution to vocabulary sophistication.

No significant main effect of native language was observed, $F(1, 18) = 0.48$, $p = .497$, $\eta^2 = .026$, suggesting a minor effect (2.6% of variation explained). A minor interaction impact was also shown by the non-significant interaction between idiom inclusion and native language ($F(1, 18) = 0.08$, $p = .781$, $\eta^2 = .004$), which only accounted for 0.4% of the variance.

When combined, the quantitative and qualitative results show a consistent pattern for both outcome variables: lexical diversity and writing quality. With large effect sizes ($\eta^2 = .24$ for writing quality and $\eta^2 = .23$ for lexical diversity), idiom inclusion had a strong and substantial impact in both situations, underscoring its beneficial contribution to perceptions of advanced language use. On the other hand, rater L1 had very little impact on either measure, and there was also little interaction between rater L1 and idiom inclusion. These findings imply that the existence of idioms was consistently seen as an indication of higher-quality, more lexically diverse writing, irrespective of whether the raters were native speakers of Hebrew or Arabic. The convergence between quantitative data and teacher commentary reinforces the role of idiomaticity as a valued and recognizable feature in L2 writing assessment.

Rater reliability analysis and challenges in evaluating idioms

While the previous sections highlighted the positive influence of idiom use, a closer look at rater reliability reveals important challenges in how consistently idioms are evaluated. For both writing quality and lexical variety ratings, Cronbach's alpha (α) showed strong internal consistency, surpassing $\alpha \geq .80$. Lexical diversity evaluations were consistent between essays with and without idioms ($\alpha = .940$ and $\alpha = .925$), while writing quality ratings were consistent between essays with and without idioms ($\alpha = .920$ and .897). Inter-rater agreement was variable, as indicated by Kendall's Tau (τ). Essays with idioms showed weak agreement ($\tau = .140$) with writing quality assessments, while essays without idioms showed moderate agreement ($\tau = .439$). Low to moderate agreement was also observed in lexical variety assessments, with $\tau = .317$ for essays that contained idioms and $\tau = .457$ for those that did not. The items in each rating set are highly connected and consistently measure the same underlying construct (writing quality and lexical diversity), as indicated by the excellent internal consistency. Idioms, on the other hand, were more difficult to consistently grade because of their subjective interpretation, which may have contributed to their lower inter-rater agreement. The fact that essays absent of idioms demonstrated greater inter-rater agreement indicates that raters can more easily evaluate straightforward language consistently.

These reliability patterns were echoed in the interviews, which shed light on the nuanced difficulties the teachers face when evaluating idioms in L2 writing. Thematic analysis revealed five key areas of concern: (1) Contextual Appropriateness, (2) Idiom Complexity (Transparency), (3) Idiom Frequency, (4) Grammatical/Syntactic Accuracy, and (5) Normative vs. Colloquial Variation. Contextual appropriateness of idioms was a major concern; improper use, such as informal idioms in academic texts or incorrect figurative meanings, frequently led to lower scores. Teacher 1 said, "You need to check whether the use is in the correct context." Teacher 5 compared the misuse of idioms to "An own goal in football... harmful," and Teacher 4 emphasized the importance of determining "Whether the student integrates the idiom in the right sense, in the right context."

Idiom complexity presented a problem since opaque idioms were harder to evaluate than transparent ones. Examples of different idiom transparency were given by Teacher 6, who pointed out that while more opaque expressions like *havle shena* ("bonds of sleep") presented more interpretive challenges and made consistent evaluation more difficult, idioms like *korea lev* ("heart-rending") were perceived as quite transparent. Additionally, the teachers pointed out that frequent idioms, although appropriate at lower proficiency levels, may indicate a lack of lexical variety in more advanced writing. In reference to idioms like *lerot et hatsi hakos hamelea* ("to see the full half of the cup"), Teacher 6 noted that "There are some idioms that have become common and frequent... they can be combined with simple writing."

Furthermore, idioms were deemed less real when they were accompanied by grammatical or syntactic errors, which frequently resulted in the assumption of memory rather than actual mastery. Teacher 2 said, "It is like a shattered or broken balloon... more harmful than helpful." Teacher 6 echoed this worry: "I would rather everything be correct without idioms than use idioms incorrectly."

Lastly, teachers' reactions to nonstandard idiomatic forms varied; some discouraged deviance from textbook usage, while others tolerated colloquial

variations as long as they matched native-speaker usage. For instance, Teacher 6 mentioned, "Even if idioms are used incorrectly, I raise the score if the incorrect form is common among native speakers." This is in reference to an idiom such as *asa yamim kelil*ot, which means "he made days as nights," rather than the more standard *asa lilot keyamim*, which means "he made nights as days" (to work day).

Discussion

Impact of idioms on writing quality and lexical diversity

The first research question looked at how native Arabic-speaking and Hebrew-speaking teachers evaluated the impact of idioms on lexical diversity and writing quality in L2 Hebrew essays. The quantitative data showed that essays with idioms got far higher scores in both writing quality ($\eta^2 = .24$) and lexical diversity ($\eta^2 = .23$), which suggests that idiomatic language has a big effect on evaluative judgments. These results are in line with previous research that found a link between idiomatic knowledge and L2 proficiency in college students (Al-kadi, 2015; Qadr & Madha, 2020). This perspective aligns with foundational theoretical frameworks, like the Idiom Principle (Sinclair, 1991) and the Lexical Approach (Lewis, 1993), which both emphasize the crucial role of formulaic language in achieving natural, fluent communication. The interview data backed up this belief even more. The teachers constantly said that being able to utilize idioms correctly and in the right context was a sign of advanced language competency. They also indicated that being able to utilize idioms correctly and in the right context was a sign of high language proficiency. For instance, Teacher 2 described idioms as "sparkling," saying they impart a text more expressive and rhetorical punch when compared to their non-idiomatic counterparts.

In line with our findings, the teachers emphasized that correct idiom use enhanced writing style, making texts sound more natural and rhetorically effective. This echoes Fernando's (1996) claim that idioms add naturalness to writing and Vasiljevic's (2015) observation that idioms contribute to vocabulary sophistication. At the same time, our interviews showed that idioms were viewed as difficult to master and easily misapplied, particularly opaque expressions, which supports prior research noting that idioms are among the most challenging aspects of L2 acquisition (Ta'amneh, 2021; Liu et al., 2019). The fact that the teachers in this study tied successful idiom use to contextual appropriateness also aligns with Atai and Akbarian's (2003) argument that exposure across varied contexts is critical for idiom acquisition. Our results therefore highlight both the stylistic rewards and the risks associated with idioms: while their correct use can elevate writing, their misuse undermines perceived competence.

The significant effect of idioms on writing quality in this study diverges from studies that suggest that the relationship between idiomatic knowledge and language ability is only modest and that idiomatic knowledge is more strongly correlated with speaking skills than with reading or writing skills (Vanderniet, 2015). This difference may stem from different methodologies. Vanderniet's (2015) study used a multiple-choice test to assess passive idiom knowledge, whereas this study examined the effect of active use of idioms in writing on teacher evaluations. The current study's findings highlight that the presence of an idiom can act as a salient marker of advanced proficiency, a contextual factor not captured by the cognitive assessment in the other study.

When it came to lexical diversity, most teachers agreed that idioms could suggest a broader vocabulary without necessarily making it richer. This distinction hints at a more complex understanding of what lexical diversity really is—that idioms can add variety to a text without necessarily introducing more difficult or less common words. The concept of lexical sophistication, i.e., the use of less frequent words in writing, is further supported by the idea that lexical diversity can be communicated through simple, frequently used vocabulary rather than advanced, infrequent words (Abu-Rabiah, 2024).

The study also explored whether a teacher's native language, Arabic or Hebrew, might shape how they assess writing quality and lexical diversity. Apparently, it did not. Native tongue appeared to have little, if any, influence on the ratings. Meaning that both native and non-native teachers perceive correct idiom usage in L2 writing as a marker of high writing quality and varied vocabularies. This finding falls in line with earlier research, which reported no significant differences between native-speaking and non-native-speaking assessors (O'Loughlin, 1994; Shi, 2001; Zhang & Elder, 2011).

Although the overall judgments were aligned, the consistency of these ratings fluctuated. The diminished inter-rater agreement (Kendall's Tau) for essays including idioms indicates that subjectivity is introduced in the assessment of idiomatic terms. This subjectivity may stem from individual differences in raters' judgments of idiomatic usage, influenced by subjective traits that affect idiom knowledge and processing, including imageability, transparency, and familiarity (Hubers et al., 2019). Moreover, idiom variation might be perceived because of linguistic subjectivity, mirroring the speaker's perspective and emotional condition (Liu, 2012). These nuances may result in variations in raters' interpretations and valuations of idioms, complicating consistent assessment. Next, the intricate decisions the teachers face when evaluating idioms in L2 student writing are discussed, offering a detailed exploration of the diminished inter-rater agreement for essays including idioms.

Taken together, the findings confirm that idioms consistently raised ratings of writing quality and lexical diversity, regardless of raters' L1. This extends earlier research (Al-kadi, 2015; Qadr & Madha, 2020) by showing that the *active use* of idioms in authentic writing tasks, not just knowledge of them, serves as a salient marker of advanced proficiency. The divergence from Vanderniet's (2015) results, where idiomatic knowledge was only modestly linked to writing ability, likely stems from methodological differences. This suggests that idioms function not only as cognitive knowledge but also as visible rhetorical signals that the teachers actively value in assessment.

Challenges in assessing idiomatic usage in L2 writing

The second research question focused on the practical and interpretive obstacles the teachers encounter when assessing idioms in student writing, highlighting that raters could more consistently assess essays without idioms. Although idioms were generally regarded as indicative of advanced writing quality and a more varied vocabulary, the qualitative data revealed that several factors influenced the variability in raters' assessments.

A significant obstacle is the contextual suitability of an idiom's usage. Educators frequently sanctioned idioms employed inappropriately, such as

incorporating informal expressions into scholarly writing or misusing figurative meanings. This underscored the imperative of both semantic and pragmatic precision. The concern is justified, as idioms exhibit significant variation in formality (Liu, 2017; Park & Chon, 2019), necessitating that learners exercise meticulous stylistic discretion. Research consistently demonstrates that context significantly influences idiom comprehension, with supportive circumstances enhancing a student's knowledge (Milburn et al., 2021; Türker, 2018). Nonetheless, due to their intrinsically figurative nature, L2 learners may still encounter difficulties with idioms, even when contextual clues are readily available (Milburn et al., 2021).

The inherent complexity of idiomatic expressions, especially opaque ones whose meanings can't be deduced from their individual parts (Wagner, 2020), was another key element. The teachers noted that evaluating these idioms required a deeper level of language and cultural savvy. This aligns with Howarth's (1998) distinction between pure idioms, which are completely opaque and often impossible to understand without prior exposure (e.g., *red herring*), and figurative idioms, whose meanings are more transparent and can be inferred from their literal components (e.g., *ring a bell*). It has been found that both production and comprehension are affected by semantic opacity; the opaquer an idiom, the harder it is to learn (Liu & Cheung, 2014). Similarly, Le Sourn-Bissaoui et al. (2012) discovered that while decomposable ambiguous idioms depend on linguistic skills and conversational perspective-taking in young people, non-decomposable ones rely more heavily on general language proficiency alone. Furthermore, Tiv et al. (2016) emphasize that transparency is what truly influences how well a person learns idioms.

Another significant concern that surfaced was the frequency of idiom usage. While appropriate for beginner- or intermediate-level writing, high-frequency idioms were sometimes considered as a sign of a limited vocabulary when they were overused in texts written at a more advanced level. This suggests that raters' expectations for idiomatic variety shift based on a student's perceived proficiency. Research supports this, with studies showing that frequency is a major driver in L2 learners' idiom acquisition. High frequency in a learner's native language can even help them learn equivalent idioms in the L2 (Türker, 2018). Moreover, when taught in meaningful situations, frequency has been shown to boost performance (Suñer, 2019), illustrating the need for contextual exposure in both L2 teaching and assessment.

Another hurdle emerged when idioms appeared alongside other linguistic errors (more complex errors, such as syntactic and semantic ones, than spelling mistakes). The teachers often questioned the authenticity of the idioms in these cases, believing them to be the result of rote memorization rather than genuine proficiency. This led to a more critical assessment overall, demonstrating a kind of negative halo effect (O'Grady, 2023), where other errors diminish the perceived value of idiomatic use. This finding echoes the findings of Vögelin et al. (2021), who found that spelling mistakes could negatively influence a teacher's judgment of a student's vocabulary. Ultimately, it seems that surface-level errors can have a disproportionate impact on the assessment of deeper lexical knowledge, like the use of idioms.

Finally, there was a clear disparity in how the teachers reacted to deviations in idiom usage. Some were sticklers for textbook standards, penalizing any departure from the norm. Others were more lenient, accepting colloquial or nonstandard versions as long as they reflected genuine native speaker use. This tension between communicative effectiveness and prescriptive accuracy mirrors a larger debate in second-language acquisition about the validity of variation and native-speaker benchmarks (Ellis et al., 2008). It also appears that a learner's proficiency level played a role in these judgments. Higher-proficiency learners are better at suppressing inappropriate figurative continuations, which allows them to produce more native-like idiomatic language (Milburn et al., 2021). The teachers might have seen some deviations not as mistakes but as evidence of advanced pragmatic ability, accepting or even encouraging them when they aligned with colloquial norms. This subjectivity, however, only underscores the challenge of applying a single set of criteria to students with varying levels of proficiency.

By going beyond the cognitive processing and receptive understanding of idioms to investigate their practical application in real-world L2 writing assessments, this study makes a genuine and unique contribution to the field. This shifts the focus from the acquisition of idioms from the perspective of a language learner to the impact of idioms on L2 assessment from the perspectives of native and non-native evaluators. Integrating both quantitative data and qualitative insights from teacher interviews provided a nuanced understanding of how idioms influence teacher perceptions of writing quality and lexical diversity, rather than just their cognitive impact.

The finding that an idiom's inclusion consistently signals higher writing quality and lexical diversity, irrespective of the rater's native language, emphasizes that idioms are indicators of high writing quality and varied vocabulary. It also demonstrates the role of idiomaticity as a valued feature in L2 writing. In addition, the finding that inter-rater agreement was lower for essays with idioms compared to those without indicates that idioms add more subjective interpretation to assessing writing quality and lexical diversity in L2 writing and complicate it. Furthermore, the identification of specific, recurring challenges the teachers face when evaluating idioms, such as contextual appropriateness, frequency, semantic opacity, departures from standard idiom usage, and co-occurrence with grammatical errors, provides a practical framework for future research and pedagogical development aimed at improving L2 writing instruction and assessment.

Conclusion

This study was set out to understand how idioms influence the evaluation of second-language writing, specifically looking at perceived lexical diversity and overall writing quality. A mixed-methods approach, which included both native Hebrew-speaking teachers and native Arabic-speaking Hebrew teachers, was used to capture not only the patterns in the final scores but also the intricate reasons behind them. The quantitative results were clear: incorporating idioms significantly increased ratings for both lexical diversity and writing quality. These findings underscore the fact that idiomatic expressions serve as a genuine sign of advanced L2 proficiency. In addition, there were no significant statistical differences in ratings between the two groups of teachers. This suggests that regardless of their

native language, both groups place a similar value on idiomatic language when evaluating L2 texts. At the same time, the qualitative data brought to light a key inconsistency: inter-rater agreement was noticeably poorer for essays with idioms, revealing a degree of subjectivity in how the teachers interpreted and judged them. The reasons for this disparity were manifold, including contextual appropriateness, frequency, semantic opacity, departures from standard idiom usage, and co-occurrence with grammatical errors. The teachers emphasized that effective idiom use requires more than knowing a phrase; it depends on contextual appropriateness, sensitivity to idiom transparency, balanced frequency, grammatical and syntactic accuracy, and awareness of standard versus colloquial variations. Ultimately, the challenges the teachers encountered while evaluating idioms emphasize the difficulty of assessing phraseological proficiency in L2 writing. While idioms can be a clear sign of high-level language skills (high writing quality and more varied vocabularies), they also introduce ambiguities that complicate the writing quality and lexical diversity assessment process.

While this study offers valuable insights into how idiomatic expressions affect the evaluation of L2 writing, its scope was limited to argumentative essays for a specific college entrance exam. This narrow focus may reduce the applicability of the findings to other genres or educational contexts. Future research should therefore explore a wider variety of writing assignments, including reflective, descriptive, and narrative essays, across a broader range of institutional and educational settings to see if these same patterns hold true.

The findings of this study lead to several practical recommendations. First, L2 learners should be encouraged to use idioms, especially common ones, in their writing. This can positively influence how raters perceive their lexical diversity and overall writing quality, particularly in high-stakes language proficiency tests. Second, to help reduce inconsistencies in scoring, assessment rubrics should be updated to specifically include criteria for idiom accuracy, frequency, and contextual appropriateness. Finally, when teaching idioms, the focus should be not only on their meaning but also on their contextual, pragmatic and stylistic use to encourage successful and appropriate deployment by students.

Acknowledgements

This study was made possible by the kind assistance of the Hebrew language teachers, who kindly donated their time and expertise to the data collection. The conceptual framework of the study was developed with assistance from K. James Hartshorn, for which the author is very grateful.

References

Abu-Gweder, A. (2023). The challenges of writing among Arab-Bedouin students between literary Arabic and Hebrew as a second language. *Journal of Namibian Studies: History Politics Culture*, *34*, 769–787. https://doi.org/10.59670/jns.v34i.1124

Abu-Rabiah, E. (2020). Lexical measures for testing progress in Hebrew as Arab students' L2. *Journal of Language and Linguistic Studies, 16*(3), 1096–1114. https://doi.org/10.17263/jlls.803551

Abu-Rabiah, E. (2023). Evaluating L2 vocabulary development features using lexical density and lexical diversity measures. *LLT Journal: A Journal on Language and Language Teaching, 26*(1), 168–182. https://doi.org/10.24071/llt.v26i1.5841

Abu-Rabiah, E. (2024). Exploring lexical sophistication in second language: An analysis of vocabulary using a word-rating method. *Theory and Practice of Second Language Acquisition, 2*(10), 1–24. https://doi.org/10.31261/TAPSLA.15623

Abu-Rabiah, E. (2025a). Semantic and phonological effects on L2 to L1 translation: The case of cognates, false cognates, and translation equivalents. *Electronic Journal of Foreign Language Teaching, 22*(1), 25–40. https://doi.org/10.56040/eiar2212

Abu-Rabiah, E. (2025b). The productive vocabulary size of second language learners upon entry into higher education. *Theory and Practice of Second Language Acquisition, 11*(1), 1–19. https://doi.org/10.31261/TAPSLA.16594

Abu-Rabiah, E. (2025c). Navigating language barriers: Hebrew proficiency and self-efficacy among Muslim Arab Bedouin students in Israeli higher education. *Journal of International Students, 15*(3), 71–90. https://doi.org/10.32674/s08gad96

Abu-Rabiah, E., Gafter, R., & Henkin, R. (2023). Where syntactic interference persists: The case of Hebrew written by native Arabic speakers. *L1-Educational Studies in Language and Literature, 23*, 1–19. https://doi.org/10.21248/l1esll.2023.23.1.399

Adams, W. C. (2015). Conducting semi-structured interviews. In K. E. Newcomer, H. P. Hatry, J. S. Wholey, K. E. Newcomer, H. P. Hatry, & J. S. Wholey (Eds.), *Handbook of practical program evaluation* (4th ed.) (pp. 492–505). John Wiley & Sons, Inc. https://doi.org/10.1002/9781119171386.ch19

Al-kadi, A. M. T. (2015). Towards idiomatic competence of Yemeni EFL undergraduates. *Journal of Language Teaching and Research, 6*(3), 513–523. http://dx.doi.org/10.17507/jltr.0603.06

Anjarini, R. D., & Hatmanto, E. D. (2021, December). Challenges and strategies in understanding English idioms: English as a foreign language students' perception. In *International Conference on Sustainable Innovation Track Humanities Education and Social Sciences* (ICSIHESS 2021) (pp. 241–247). Atlantis Press. http://dx.doi.org/10.2991/assehr.k.211227.039

Atai, M. R., & Akbarian, I. (2003). The effect of exposure on EFL learners' acquisition of idioms with reference to proficiency levels. *Indian Journal of Applied Linguistics, 29*(1), 21–34.

Bestgen, Y., & Granger, S. (2014). Quantifying the development of phraseological competence in L2 English writing: An automated approach. *Journal of Second Language Writing, 26*, 28–41. https://doi.org/10.1016/j.jslw.2014.09.004

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Cacciari, C. (2014). Processing multiword idiomatic strings: Many words in one?. *The Mental Lexicon, 9*(2), 267–293. https://doi.org/10.1075/ml.9.2.05cac

Cacciari, C., & Tabossi, P. (2014). *Idioms: Processing, structure, and interpretation*. Psychology Press.

Carrol, G., & Conklin, K. (2020). Is all formulaic language created equal? Unpacking the processing advantage for different types of formulaic sequences. *Language and Speech, 36*(1), 95–122. https://doi.org/10.1177/0023830918823230

Casas, R. and Campoy, J. (1995). A sociolinguistic approach to the study of idioms: Some anthropolinguistic sketches. *Cuadernos de Filología Inglesa, 4*, 43–61.

Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research, 11*(3), 415–443. https://doi.org/10.17239/jowr-2020.11.03.01

Crossley, S. A., & McNamara, D. S. (2012). Predicting second language writing proficiency: The roles of cohesion and linguistic sophistication. *Journal of Research in Reading, 35*(2), 115–135. https://doi.org/10.1111/j.1467-9817.2010.01449.x

Ellis, N. C., Simpson-Vlach, R. I. T. A., & Maynard, C. (2008). Formulaic language in native and second language speakers: Psycholinguistics, corpus linguistics, and TESOL. *TESOL Quarterly, 42*(3), 375–396. https://doi.org/10.1002/j.1545-7249.2008.tb00137.x

Espinal, M. T., & Mateu, J. (2019). Idioms and phraseology. In M. T. Espinal & J. Mateu (Eds.), *Oxford research encyclopedia of linguistics*. Oxford University Press. https://doi.org/10.1093/acrefore/9780199384655.013.51

Gramley, S., Gramley, V., & Pätzold, K. M. (2020). *A survey of modern English*. Routledge. https://doi.org/10.4324/9780429300356

Hajiyeva, B. (2024). Challenges in understanding idiomatic expressions. *Acta Globalis Humanitatis et Linguarum, 1*(2), 67–73. https://doi.org/10.69760/aghel.01024066

Haskel-Shaham, I., Klaus, R., & Tamir, R. (2018). Discourse—yes, grammar— no. Influence of Arabic mother tongue on Arab students writing in Hebrew. *L1-Educational Studies in Language and Literature, 18*, 1–26. https://doi.org/10.17239/L1ESLL-2018.18.01.07

Henkin, R. (2020). Persistence of interference from L1 Arabic in written Hebrew. *L1-Educational Studies in Language and Literature, 20*, 1–31. https://doi.org/10.17239/L1ESLL-2020.20.01.15

Howarth, P. (1998). The phraseology of learners' academic writing. In A. Cowie (Ed.), *Phraseology: Theory, analysis, and applications* (pp. 161–186). Oxford University Press. https://doi.org/10.1093/oso/9780198294252.003.008

Hubers, F., Cucchiarini, C., Strik, H., & Dijkstra, T. (2019). Normative data of Dutch idiomatic expressions: Subjective judgments you can bank on. *Frontiers in Psychology, 10*, 1075. https://doi.org/10.3389/fpsyg.2019.01075

Jarvis, S. (2017). Grounding lexical diversity in human judgments. *Language Testing, 34*(4), 537–553. https://doi.org/10.1177/0265532217710632

Johnson, R. B., Onwuegbuzie, A. J., & Turner, L. A. (2007). Toward a definition of mixed methods research. *Journal of Mixed Methods Research, 1*(2), 112–133. https://doi.org/10.1177/1558689806298224

Joshi, A., Kale, S., Chandel, S., & Pal, D. K. (2015). Likert scale: Explored and explained. *British Journal of Applied Science & Technology, 7*(4), 396–403. https://doi.org/10.9734/BJAST/2015/14975

Kashiha, H., & Chan, S. H. (2015). A little bit about: Differences in native and non-native speakers' use of formulaic language. *Australian Journal of Linguistics, 35*(4), 297–310. https://doi.org/10.1080/07268602.2015.1067132

Kiger, M. E., & Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical Teacher, 42*(8), 846–854. https://doi.org/10.1080/0142159X.2020.1755030

Le Sourn-Bissaoui, S., Caillies, S., Bernard, S., Deleau, M., & Brulé, L. (2012). Children's understanding of ambiguous idioms and conversational perspective-taking. *Journal of Experimental Child Psychology, 112*(4), 437–451. https://doi.org/10.1016/j.jecp.2012.02.003

Lewis, M. (1993). *The lexical approach: The state of ELT and a way forward*. Language Teaching Publications.

Li, L. (2023). *Quadri-syllabic schematic idioms in Chinese: Description and acquisition*. Springer. https://doi.org/10.1007/978-981-19-7202-7

Listyani, L., & Thren, A. T. (2023). Indonesian EFL learners' errors in idiomatic expressions: Teachers' and students' perspectives. *European Journal of Education Studies, 10*(10), 374–392. http://dx.doi.org/10.46827/ejes.v10i10.5036

Liu, D. (2017). *Idioms: Description, comprehension, acquisition, and pedagogy*. Routledge. https://doi.org/10.4324/9781315092843

Liu, L., & Cheung, H. T. (2014). Acquisition of Chinese quadra-syllabic idiomatic expressions: Effects of semantic opacity and structural symmetry. *First Language, 34*(4), 336–353. https://doi.org/10.1177/0142723714544409

Liu, Y., Pang, B., & Liu, B. (2019). Neural-based Chinese idiom recommendation for enhancing elegance in essay writing. *Annual Meeting of the Association for Computational Linguistics*.

Liu, Z. (2012). Analysis of idiom variation in the framework of linguistic subjectivity. *English Language Teaching, 5*(6), 105–113.

Malvern, D., Richards, B., Chipere, N., and Durán, P. (2004). *Lexical diversity and language development: Quantification and assessment*. Palgrave Macmillan. https://doi.org/10.1057/9780230511804

McCarthy, P. M., & Jarvis, S. (2007). Vocd: A theoretical and empirical evaluation. *Language Testing, 24*(4), 459–488. https://doi.org/10.1177/0265532207080767

McGlone, M. S., Glucksberg, S., & Cacciari, C. (1994). Semantic productivity and idiom comprehension. *Discourse Processes, 17*(2), 167–190. https://doi.org/10.1080/01638539409544865

Meara, P. (2014). Review of the book Vocabulary knowledge: Human ratings and automated measures, by Scott Jarvis and Michael Daller (Eds.). *International Journal of Applied Linguistics, 24*(3), 418–421. https://doi.org/10.1111/ijal.12086

Mehdi, S. (2024). Techniques used in teaching idioms. *Acta Globalis Humanitatis et Linguarum, 1*(2), 34–42. https://doi.org/10.69760/aghel.01024063

Milburn, E., Vulchanova, M., & Vulchanov, V. (2021). Collocational frequency and context effects on idiom processing in advanced L2 speakers. *Canadian Journal of Experimental Psychology, 75*(2), 169–174. https://doi.org/10.1037/cep0000214

Miller, J. (2020). The bottom line: Are idioms used in English academic speech and writing? *Journal of English for Academic Purposes, 43*, Article 100810. https://doi.org/10.1016/j.jeap.2019.100810

O'Grady, S. (2023). Halo effects in rating data: Assessing speech fluency. *Research Methods in Applied Linguistics, 2*(2), Article 100048. https://doi.org/10.1016/j.rmal.2023.100048

O'Loughlin, K. J. (1994). The assessment of writing by English and ESL teachers. *Australian Review of Applied Linguistics, 17*(1), 23–44. https://doi.org/10.1075/aral.17.1.02olo

Paquot, M. (2018). Phraseological competence: A missing component in university entrance language tests? Insights from a study of EFL learners' use of statistical collocations. *Language Assessment Quarterly, 15*(1), 29–43. https://doi.org/10.1080/15434303.2017.1405421

Park, J., & Chon, Y. V. (2019). EFL learners' knowledge of high-frequency words in the comprehension of idioms: A boost or a burden? *RELC journal, 50*(2), 219–234. https://doi.org/10.1177/00336882177480

Plano Clark, V. L. (2017). Mixed methods research. *The Journal of Positive Psychology, 12*(3), 305–306. https://doi.org/10.1080/17439760.2016.1262619

Qadr, H. M., & Madhat, I. R. (2020). The relationship between idiomatic knowledge and second language proficiency/Kurdistan universities undergraduates as an example. *International Journal of Linguistics, Literature and Translation, 3*(8), 149–160. https://al-kindipublisher.com/index.php/ijllt/article/view/374

Rafatbakhsh, E., & Ahmadi, A. (2019). A thematic corpus-based study of idioms in the Corpus of Contemporary American English. *Asian-Pacific Journal of Second and Foreign Language Education, 4*(1), Article 11. https://doi.org/10.1186/s40862-019-0076-4

Saberian, N., & Fotovatnia, Z. (2011). Idiom taxonomies and idiom comprehension: Implications for English teachers. *Theory and Practice in Language Studies, 1*(9), 1231–1235.

Senaldi, M. S. G., & Titone, D. (2024). Idiom meaning selection following a prior context: Eye movement evidence of L1 direct retrieval and L2 compositional assembly. *Discourse Processes, 61*(1–2), 21–43. https://doi.org/10.1080/0163853X.2024.2311637

Shi, L. (2001). Native- and nonnative-speaking EFL teachers' evaluation of Chinese students' English writing. *Language Testing, 18*(3), 303–325. https://doi.org/10.1177/026553220101800303

Sinclair, J. (1991). *Corpus, concordance, collocation*. Oxford University Press

Siyanova-Chanturia, A., Conklin, K., & Schmitt, N. (2011). Adding more fuel to the fire: An eye-tracking study of idiom processing by native and non-native speakers. *Second Language Research, 27*(2), 251–272. https://doi.org/10.1177/0267658310382068

Suhodolli, M., & Lama, A. (2024). Teaching English idioms: A focus on cultural adaptation and contextual learning in Kosovo—Case study: Public international business college in Mitrovica, Kosovo). *Journal of Educational and Social Research, 14*(6), 585–595. https://doi.org/10.36941/jesr-2024-0195

Suñer, F. (2019). Exploring the relationship between stay-abroad experiences, frequency effects, and context use in L2 idiom comprehension. *Journal of Language Teaching and Research, 10*(1), 35–44. http://dx.doi.org/10.17507/jltr.1001.04

Ta'amneh, M. A. A. A. (2021). Strategies and difficulties of learning English idioms among university students. *Strategies, 12*(23), 76–84.

Tiv, M., Milburn, E., & Warren, T. (2016). Investigating the effects of transparency and ambiguity on idiom learning. *Annual Meeting of the Cognitive Science Society*.

Türker, E. (2018). The influence of L1 frequency in instructed second language learning of L2 idioms. *Journal of Second Language Studies, 1*(2), 333–356. https://doi.org/10.1075/jsls.17007.tur

Vanderniet, K. H. (2015). *Idioms as a measure of proficiency*. Brigham Young University.

Vasiljevic, Z. (2015). Teaching and learning idioms in L2: From theory to practice. *MAXTESOL Journal, 39*(4), 1–24.

Vögelin, C., Jansen, T., Keller, S. D., & Möller, J. (2021). The impact of vocabulary and spelling on judgments of ESL essays: An analysis of teacher comments. *The Language Learning Journal, 49*(6), 631–647. https://doi.org/10.1080/09571736.2018.1522662

Wagner, W. (2020). *Idioms and ambiguity in context: Phrasal and compositional readings of idiomatic expressions*. Mouton De Gruyter. https://doi.org/10.1515/9783110685459

Zhang, Y., & Elder, C. (2011). Judgments of oral proficiency by non-native and native English-speaking teacher raters: Competing or complementary constructs? *Language Testing, 28*(1), 31–50. https://doi.org/10.1177/0265532209360671