

Paper Type: Original Article



Linguistic and Methodological Features of Research Questions in Mixed Methods TEFL Dissertations: A Corpus-Based Analysis

Fatemeh Zamanian^{1,*}, Amin Karimnia², Zohreh Seifoori³¹ Department of English Language, Islamic Azad University, Science and Research Branch, Tehran, Iran;² Department of English Language and Literature, Islamic Azad University, Fasa Branch, Fasa, Iran;³ Department of English Language, Islamic Azad University, Science and Research Branch, Tehran, Iran.

Received: 28 January, 2026

Revised: 03 June, 2026

Accepted: 14 June, 2026

Abstract

Research questions (RQs) play a pivotal role in mixed methods research (MMR), guiding the integration of qualitative and quantitative strands to address complex inquiries. This study investigates how the linguistic formulation of RQs reflects, rather than determines, the methodological choices made in MMR-based PhD dissertations in Teaching English as a Foreign Language (TEFL). Drawing on Creswell's (2009) mixed methods design typology, a corpus of 518 RQs extracted from 92 Iranian TEFL dissertations (2019–2024) was analyzed using a manually coded corpus-based approach complemented by qualitative content analysis. RQs were operationally defined as all explicitly numbered or labeled research questions, excluding aims, hypotheses, and embedded declarative statements. The quantitative findings show a clear preference for sequential designs, particularly explanatory (39.1%) and exploratory (34.8%), which is a pattern consistent with TEFL's emphasis on staged, context-sensitive inquiry. Regarding syntactic realization, non-polar Wh-questions were most frequent (56.8%), with factual (62.6%) and evaluative (20.4%) types prevailing across designs. Simple sentence structures (66.0%) and present-tense framing (99.2%) dominated, indicating a preference for clarity and alignment with the conventions of dissertation writing rather than discipline-wide tendencies. The qualitative findings illustrate how these syntactic patterns support transparency in reporting and help structure methodological decision-making. The findings contribute to understanding how researchers in Iran construct RQs in applied linguistics landscape and provide a foundation for cross-context comparisons.

Keywords: Research question formulation, Syntactic features, Mixed-methods research (MMR), TEFL dissertations, Genre analysis, EAP pedagogy.



Journal of Studies in Language Learning and Teaching. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY-NC) license.

I | INTRODUCTION

In Mixed Methods Research (MMR), the formulation of Research Questions (RQs) serves as the primary "directional determinant" (Feak & Swales, 2011) that shapes the scope, coherence, and methodological rigor of a study. Unlike mono-method inquiries, RQs in MMR must explicitly signal how diverse paradigms interact, functioning as the methodological "anchor" for the entire research project. Despite their importance, there is a lack of systematic research on how these questions are linguistically constructed, particularly regarding their syntactic realization (e.g., polarity, clause structure, and tense), and how these choices correlate with the overarching design of the study. This

✉ Corresponding Author: zamanianfatima62@gmail.com

Cite this article: Zamanian, F., Karimnia, A., Seifoori, Z., (2026). Linguistic and Methodological Features of Research Questions in Mixed Methods TEFL Dissertations: A Corpus-Based Analysis, *Journal of Studies in Language Learning and Teaching*, 2(2), 121-135. doi: 10.22034/JSLLT.2026.24278.1115.

study addresses that gap by shifting the focus from general genre structures to the RQ itself, treating the dissertation introduction chapter strictly as the site of data extraction where these critical methodological determinants are articulated.

The study focuses on MMR because it represents a prominent, yet rhetorically demanding, approach in TEFL research. MMR requires researchers to integrate qualitative and quantitative strands to capture complex, context-embedded educational phenomena (Creswell & Plano Clark, 2018). Formulating RQs in this context is particularly challenging, as it requires precise alignment between distinct paradigms and stages of inquiry (Onwuegbuzie & Leech, 2006). Previous scholarship has noted significant weaknesses in the explicit articulation of MMR designs in applied linguistics, with Riazi and Farsani (2024) reporting that 70.4% of published studies did not clearly specify their design typology. Understanding how doctoral writers construct RQs offers a crucial window into how early-stage researchers conceptualize and report their mixed methods studies.

While extensive genre analysis has examined the rhetorical move structures of dissertation introductions (e.g., Kawase, 2018; Ebadi et al., 2019; Soodmand Afshar & Ranjbar, 2017), the specific syntactic properties of RQs have received far less attention. Recent studies investigating RQs in applied linguistics (e.g., Ghanbar & Rezvani, 2023; Lim & Luo, 2020) highlight variation in linguistic formulation, but they do not examine how these variations manifest across specific MMR design types. As such, this study bridges genre analysis and methodological reporting by examining linguistic patterns in RQs within a corpus of Iranian TEFL doctoral dissertations.

In this study, syntactic features are defined operationally and mutually exclusively as sentence structure (simple, compound, complex, and compound-complex), polarity (polar/yes-no vs. non-polar/content questions), and verb tense (present vs. future). These features were manually coded using a detailed scheme based on standard grammatical descriptions (Azar & Hagen, 2009; Radford, 2009). Our intention is not to claim that syntactic features determine methodological success; rather, we argue that the syntactic realization of RQs tends to reflect the methodological decisions and reporting conventions adopted by doctoral researchers.

The national focus on Iranian PhD dissertations is justified for three reasons. First, Iran has an active and well-established TEFL research community, producing a substantial body of doctoral work. Second, many Iranian dissertations are publicly accessible, allowing transparent corpus construction. Third, the Iranian TEFL context offers insight into how NNS researchers position themselves in the broader landscape of applied linguistics. That said, we avoid generalizing the findings to "global TEFL" and instead describe the contribution as offering context-grounded insights that may be compared with other settings in future research.

The study is guided by the following research questions:

1. What is the distribution of Mixed-methods Research (MMR) designs employed in Iranian TEFL PhD dissertations based on Creswell's (2009) framework?
2. Which syntactic features (polarity, sentence structure, verb tense) characterize the formulation of research questions (RQs) in these dissertations?
3. How do syntactic features vary across different MMR design types, and what do these patterns reveal about reporting practices?

By addressing these questions, this study contributes to the literature in two ways. First, it provides an operationalized and replicable linguistic framework for analyzing RQ formulation in MMR. Second, it offers pedagogically grounded guidelines to support English for Academic Purposes (EAP) instruction for novice researchers in applied linguistics and TEFL. Together, these contributions situate the study at the intersection of genre analysis and methodological transparency.

II. LITERATURE REVIEW

Mixed Methods Research (MMR) developed as an approach that integrates qualitative and quantitative paradigms. Tashakkori and Teddlie (2003) argued that relying on only one paradigm limits inquiry, while Johnson et al. (2007) defined MMR as the deliberate combination of both to achieve fuller understanding. Collins et al. (2006) outlined 13 procedural phases from defining purposes to validating findings, showing the iterative and adaptable nature of MMR.



As research phenomena in social sciences have grown more complex, MMR has gained recognition. Venkatesh et al. (2013) emphasized its contribution to developing holistic models, while applications in education, health, and psychology highlight its versatility. Despite its promise, Riazi and Farsani (2024) found that 70.4% of MMR studies in applied linguistics lacked explicit design specification, pointing to weaknesses in research question (RQ) formulation. This study responds by examining the syntactic features of RQs in TEFL dissertations, emphasizing their role in ensuring methodological rigor.

2.1. MMR in TEFL research

In TEFL, MMR is valuable for examining language acquisition, pedagogy, and motivation. Dawadi et al. (2021) identified three advantages: analyzing complex social phenomena comprehensively, enabling confirmatory and explanatory inquiries simultaneously, and reconciling conflicting findings. For example, studies may combine quantitative test data with qualitative interviews to capture multiple dimensions of instruction (Creswell & Plano Clark, 2018). This study extends such work by analyzing how TEFL RQs align syntactically with MMR designs.

2.2. Research questions in MMR

RQs are central in MMR because they guide integration of methods. Merchant et al. (2021) showed that only 11% of the examined instructional design studies framed RQs according to MMR principles. Similarly, Plano Clark and Badiee (2010) and Tashakkori and Creswell (2007) observed limited attention to this issue in postgraduate contexts. Riazi (2017) argued that weak RQs undermine methodological coherence. This study focuses on how syntax, namely sentence type, polarity, and tense, shapes clarity and integration in TEFL dissertations.

While early research on RQs emphasized their rhetorical functions, such as move structures in introductions (Ozturk, 2007; Shehzad, 2011; Sheldon, 2011) and discourse roles (Soler-Monreal et al., 2011), other scholars investigated the linguistic nature of these inquiries. Lim (2014) studied linguistic mechanisms in experimental dissertations, and Riazi and Candlin (2014) surveyed RQs in journals, though neither focused on the specific syntax required by MMR. Similarly, Kawase (2018) and Lu et al. (2021) examined frequency patterns, and Ghanbar and Rezvani (2023) explored syntactic features in L2 research, but these studies did not differentiate between specific methodological designs.

Recent inquiry has further refined the understanding of RQ formulation by focusing on methodological alignment. Soodmand Afshar and Hafez (2021) emphasized the necessity of aligning RQs with the methodological paradigm to ensure internal consistency. Despite these insights, gaps remain. For example, Kross and Giust (2019) focused on qualitative RQs, and Amini Farsani and Mohammadi (2022) evaluated MMR quality by domain and scope, but they neglected the linguistic formulation. This study addresses these gaps by analyzing polarity, sentence structure, and tense to demonstrate how syntactic features reflect methodological goals and enhance transparency.

Building upon the existing scholarly interest in RQ analysis (e.g., Ghanbar & Rezvani, 2023; Kross & Giust, 2019), the present study extends this line of inquiry by systematically examining the syntactic construction of RQs specifically within TEFL mixed-methods research (MMR) dissertations. By focusing on a corpus of Iranian Ph.D. dissertations, the study illuminates how linguistic realization links to broader methodological transparency and rhetorical clarity.

The findings provide valuable pedagogical implications for English for Academic Purposes (EAP), offering both students and instructors practical strategies for crafting well-structured and effective RQs. Moreover, the study opens avenues for cross-cultural comparisons, enriches the understanding of MMR practices, and contributes to strengthening TEFL research standards and scholarly rigor at a global level.

III. METHODOLOGY

This study adopted a mixed-methods analytical approach to examine the linguistic and methodological characteristics of research questions (RQs) in Iranian PhD dissertations in Teaching English as a Foreign Language (TEFL). The section below clarifies corpus construction, coding procedures, reliability, and statistical analysis, while distinguishing between quantitative and qualitative components.

3.1. Research design

This study employed a sequential explanatory mixed-methods design, following Creswell's (2009) framework. The design consisted of two phases. In the quantitative phase, a corpus-based analysis of RQs was performed, focusing on the distribution of syntactic features across mixed-methods research (MMR) designs.

In the qualitative phase, the qualitative content was analyzed with the aim of interpreting how doctoral writers used linguistic features in ways that supported methodological reporting and rhetorical clarity. The qualitative phase contextualizes these reflections by analyzing how RQs are framed within each dissertation's methodological narrative.

3.2. Corpus and sampling

The corpus consisted of 92 PhD dissertations (2019–2024) collected from publicly accessible digital repositories of major Iranian universities. The repositories included the National Thesis and Dissertation Database (IRANDOC), Islamic Azad University Digital Repository, and the institutional repositories of state universities such as Shiraz University, Allameh Tabataba'i University, and Ferdowsi University of Mashhad.

The corpus selection process was guided by specific inclusion and exclusion criteria. The dissertations were required to be full-text accessible, written in English, and explicitly identified by the authors through mixed-methods research. Conversely, the mono-method studies, incomplete or inaccessible documents, and those lacking numbered or clearly labeled research questions were excluded. To ensure a diverse representation across institutions, supervisors, topic areas (e.g., CALL, teacher cognition, and vocabulary development), and MMR design types, maximum variation sampling was employed. Only the dissertations that had completed the formal university examination process were included in the final analysis.

3.3. Data collection

A corpus of 92 dissertations was purposefully compiled to ensure both thematic and methodological diversity, drawn from the publicly accessible repositories across 15 Iranian State and Islamic Azad universities. All the author identities were anonymized to maintain confidentiality throughout the research process. For the data collection, the RQs were extracted manually rather than through automated means, ensuring accuracy. AntConc was used only for frequency-based lexical extraction, not syntactic analysis. An RQ was defined as any explicitly numbered or labeled interrogative sentence presented under a section titled "Research Questions", "Questions of the Study", or equivalent headings. The following were excluded:

- Aim or purpose statements
- Hypotheses
- Declarative statements
- Interrogatives embedded in paragraphs

The data collection proceeded in two stages. The first was the quantitative analysis of the RQs, which were identified and extracted manually to ensure accurate syntactic coding. The AntConc software was utilized only as a supplementary tool for frequency-based lexical verification and corpus management. Following this manual extraction, syntactic features including sentence structure (simple, compound, complex, and compound-complex), polarity (polar vs. non-polar), and verb tense (present vs. future) were documented in a detailed Excel-based coding scheme aligned with established grammatical frameworks (Azar & Hagen, 2009; Radford, 2009). In the second stage, a qualitative content analysis was conducted on a subset of 20 dissertations. These were selected using purposive sampling to ensure a balanced representation across key thematic areas (e.g., vocabulary development, teacher cognition) and the three primary MMR design types identified in our initial corpus (explanatory, exploratory, and triangulation). This qualitative stage examined how specific syntactic configurations facilitated academic argumentation, utilizing illustrative examples to highlight prevalent descriptive and evaluative patterns. Ethical standards were strictly maintained, as all the dissertations were publicly available through institutional repositories. No identifiable personal data were collected. Only RQs were extracted and analyzed. The study was granted exemption status by the researchers' institutional ethics committee because it used publicly accessible documents for secondary analysis. The efforts to ensure anonymity included avoiding the direct quotation of the long RQs that could reveal authors, reporting aggregated patterns rather than individual-level details. The data were stored securely on an encrypted drive; they will not be shared publicly, in line with university data-management guidelines.

3.4. Data analysis

The analysis proceeded in several phases to address the research questions systematically as follows:

1. MMR design analysis: The MMR designs were classified using Creswell's (2009) framework into concurrent (Triangulation, Embedded, Transformative) and sequential designs (Explanatory, Exploratory, Transformative). This classification ensured consistency with prior TEFL studies (Amini Farsani & Mohammadi, 2022) and provided a structured basis for analyzing the RQ distribution across the designs.

2. Coding scheme for syntactic features: The RQs were analyzed for syntactic features across the following dimensions. A detailed coding scheme was developed based on descriptive grammar sources (Azar & Hagen, 2009; Radford, 2009). All the syntactic features were coded manually.

a. Sentence structure (mutually exclusive):

- Simple: One independent clause
- Complex: One independent clause + one or more dependent clauses
- Compound: Two or more independent clauses joined by coordinating conjunctions
- Compound-complex: Two or more independent clauses + at least one dependent clause

Of note here is the scarcity of RQs with a compound-complex sentence structure. Such cases did not occur in the corpus.

b. Polarity:

- Polar: Yes/no questions
- Non-polar: Wh-questions

c. Wh-question type:

- Factual: Requesting factual/descriptive information
- Evaluative: Asking about value, quality, or extent
- Circumstantial: Addressing conditions, contexts, or circumstances
- Determinative: Seeking clarification of amount, scope, or boundaries
- Causative: Focusing on reasons or underlying causes
- Personal Wh-type: Inquiring about human agents, roles, or participants

Causative and personal Wh-types were included in the original coding scheme, but they did not appear in the corpus.

d. Verb tense:

- Present (e.g., "What factors influence...?")
- Future (e.g., "What will be the effect...?")

A detailed coding protocol was created to ensure consistency and avoid overlap among the categories.

3. Qualitative content analysis: A qualitative content analysis was conducted on a purposive subsample of 20 dissertations selected to represent all the MMR design types. The analysis examined how the research questions (RQs) were introduced and justified, how their linguistic form related to stage-based methodological rationales, and how doctoral writers connected RQ wording to sampling procedures, instrumentation, and analytical decisions. A qualitative content analysis was conducted on a purposive subsample of 20 dissertations selected to represent all MMR design types. The analysis examined how research questions (RQs) were introduced and justified, how their linguistic form related to stage-based methodological rationales, and how doctoral writers connected RQ wording to sampling procedures, instrumentation, and analytical decisions. Coding was performed in NVivo using an inductive thematic approach with a primary focus on the rhetorical function of RQs. This qualitative phase was designed to illuminate the rhetorical work that RQs perform within doctoral mixed-methods dissertations, with the resulting thematic patterns integrated into the subsequent results and discussion sections to provide a more holistic understanding of the data.

4. Coding procedure and inter-coder reliability: Two trained coders carried out the coding process. Both coders held graduate-level qualifications in applied linguistics and had prior experience in discourse and textual analysis. Before beginning the analysis, they completed a three-hour training session that covered the identification of the research questions, the classification of the mixed-methods designs based on Creswell's (2009) framework, and the application of the syntactic coding scheme developed for this study. To assess inter-coder reliability, 20% of the dataset (104 RQs) was independently coded by both coders. The resulting Cohen's Kappa coefficients demonstrated strong reliability across all the categories: sentence structure ($\kappa = 0.82$), polarity ($\kappa = 0.91$), Wh-type ($\kappa = 0.79$), verb tense ($\kappa = 0.94$), and MMR design classification ($\kappa = 0.86$). These values fall within the "substantial" to "almost perfect" range according to

Altman (1991). Any discrepancies were discussed collaboratively, and the coding scheme was refined before the remaining corpus was analyzed.

5. Statistical analysis: Descriptive statistics (means, frequencies, percentages) were analyzed using IBM SPSS Statistics v22. For the MMR design distribution, frequencies and percentages were computed. For the syntactic features, means and frequencies of sentence structures, question forms, and verb tenses were calculated. For consistency across the MMR designs, Chi-square tests assessed the significant distributional differences. The quantitative and qualitative findings were integrated through meta-inferences, where quantitative patterns (e.g., dominance of non-polar questions) were contextualized by qualitative themes (e.g., emphasis on descriptive inquiries), ensuring a cohesive analysis.

IV. RESULTS

This section presents the findings addressing the three research questions, which explore the distribution of mixed methods research (MMR) designs, the syntactic features of research questions (RQs), and their consistency across MMR designs in PhD TEFL dissertations. The analysis integrates descriptive statistics (frequencies, means, and percentages) with inferential tests (Chi-square) to provide a comprehensive overview. To enhance interpretive depth, the findings are contextualized in TEFL disciplinary norms and compared with the MMR trends in related fields (e.g., education, psychology). Figure 1 illustrates the relationship between RQ syntactic features and MMR designs, offering a visual summary of key patterns.

4.1. Distribution of MMR designs in Ph.D. TEFL dissertations

The distribution of MMR designs, classified using Creswell's (2009) framework, reveals a clear preference for sequential approaches in TEFL dissertations. As shown in Table 1, sequential designs dominate, accounting for 73.9% ($n = 68$) of the 92 dissertations, with explanatory designs being the most prevalent (39.1%, $n = 36$), followed by exploratory designs (34.8%, $n = 32$). Concurrent designs, comprising triangulation (21.7%, $n = 20$) and embedded (4.3%, $n = 4$) designs, are less common. The absence of transformative designs suggests a focus on practical, context-driven inquiries over theoretical paradigm shifts.

This preference for sequential designs reflects TEFL's emphasis on phased investigations, where quantitative findings often inform qualitative exploration (e.g., test scores followed by interviews) or qualitative insights guide instrument development. The sequential dominance of TEFL may stem from the need to address complex, context-specific phenomena, such as language acquisition in Iran's EFL settings, which require iterative data collection to capture cultural nuances.

Table 1. Distribution of MMR designs in PhD TEFL dissertations

MMR design category		Frequency	Percent
Sequential	Explanatory	36	39.1
	Exploratory	32	34.8
	Triangulation	20	21.7
Concurrent			
	Embedded	4	4.3
Total		92	100.0

4.2. RQ distribution and word count means across MMR designs

Table 2 details the distribution of RQs and their mean word counts across MMR designs. The sequential explanatory design accounts for the highest proportion of dissertations (39.1%, $n = 36$) and RQs (42.1%, $n = 218$), with a mean of 6.05 RQs per dissertation and a mean word count of 114.5. This suggests that explanatory designs prioritize comprehensive inquiries, often addressing multiple facets of pedagogical interventions. The sequential exploratory design follows, with 34.7% of dissertations ($n = 32$) and 28.6% of RQs ($n = 148$), averaging 4.62 RQs and a concise mean word count of 91.1, indicating focused, instrument-driven studies. In contrast to these sequential approaches, where one phase informs the next, the concurrent designs, triangulation and embedded, collect data simultaneously. Triangulation designs (21.7%, $n = 20$; 22.0% RQs, $n = 114$) maintain moderate RQ counts ($M = 5.7$) and word counts ($M =$



112.9), reflecting balanced integration. Embedded designs, though rare (4.3%, $n = 4$), have the highest mean RQ count (9.5) and word count (263.0), suggesting complex, in-depth inquiries often requiring multifaceted concurrent data streams.

The high RQ count in embedded designs may reflect their use in experimental studies, where qualitative components elucidate quantitative outcomes. In contrast, TEFL preference for concise exploratory RQs aligns with its focus on practical, context-sensitive research, particularly in Iran, where cultural and institutional factors emphasize actionable findings for educators.

Table 2. RQ distribution and word count means across MMR designs

MMR design	Total dissertations (%)	Total RQs (%)	Mean RQs	Mean word count
sequential				
Explanatory	36 (39.1)	218 (42.1)	6.05	114.5
Exploratory	32 (34.7)	148 (28.6)	4.62	91.1
Concurrent				
Triangulation	20 (21.7)	114 (22.0)	5.7	112.9
Embedded	4 (4.3)	38 (7.3)	9.5	263.0
Total	92 (100.0)	518 (100.0)	5.63	111.2

Note: Total Dissertations = Total Number of Dissertations; Total RQs = Total Number of Research Questions; Mean RQs = Mean Number of Research Questions; Mean Word Count = Mean Number of Words per RQ.

4.3. Polar and non-polar RQ distribution across MMR designs

A Chi-square test of independence revealed significant differences in the distribution of polar and non-polar RQs across the MMR designs, $\chi^2 (3, N = 518) = 39.48, p < .001$, with a moderate effect size (Cramér's $V = 0.276$). As shown in Table 3, non-polar Wh-questions dominate (56.8%, $n=294$), with sequential explanatory designs contributing the largest share (44.2%, $n = 130$), followed by exploratory designs (33.3%, $n = 98$). Polar RQs, which elicit yes/no responses, are less common (43.2%, $n = 224$), with explanatory designs again leading (39.2%, $n = 88$). The embedded designs show the lowest proportion of non-polar RQs (1.3%, $n = 4$), likely due to their limited use in the corpus.

The dominance of non-polar Wh-questions reflects the exploratory orientation of TEFL, prioritizing descriptive and evaluative inquiries (e.g., "What is the effect of web-based DA on grammatical accuracy?") over confirmatory ones. The preference for non-polar questions may also stem from Iran's EFL context, where researchers aim to address practitioner-oriented questions that inform classroom practices.

Table 3. Polar and non-polar RQs distribution across MMR designs

MMR design	Polar RQs (%)	Mean polar RQs	Non-polar RQs (%)	Mean non-polar RQs
sequential				
Explanatory	88 (39.2)	2.4	130 (44.2)	3.6
Exploratory	50 (22.3)	1.5	98 (33.3)	3.0
Concurrent				
Triangulation	52 (23.2)	2.6	62 (21.1)	3.1
Embedded	34 (15.1)	8.5	4 (1.3)	1.0
Total	224 (100.0)		294 (100.0)	

Note: Polar RQs = Total Number of Polar Research Questions; Mean Polar RQs = Mean Number of Polar Research Questions; Non-Polar RQs = Total Number of Non-Polar Research Questions; Mean Non-Polar RQs = Mean Number of Non-Polar Research Questions.

4.4. Distribution of Wh-question types across MMR designs

The distribution of Wh-question types is detailed in Table 4. Factual Wh-questions are the most prevalent (62.6%, $n = 184$), followed by evaluative (20.4%, $n = 60$) and circumstantial (10.9%, $n = 32$) questions,

while determinative Wh-questions are rare (6.1%, $n = 18$). Notably, causative and personal Wh-questions were entirely absent from the corpus, indicating that doctoral researchers in this context do not utilize these specific interrogative types when formulating their RQs.

The absence of causative and personal Wh-questions suggests a gap in TEFL research, potentially due to the preference of the field for descriptive over causal inquiries or a focus on group-level rather than individual learner perspectives. The emphasis on factual and evaluative questions aligns with Iran's practitioner-oriented research culture, aiming to inform curriculum design and teaching practices.

Table 4. Distribution of Wh-question types across MMR designs

MMR Design	Factual	Circumstantial	Evaluative	Determinative	Causative	Personal
Sequential						
Explanatory	80 (27.2%)	16 (5.4%)	26 (8.8%)	8 (2.7%)	0	0
Exploratory	64 (21.8%)	8 (2.7%)	20 (6.8%)	6 (2.0%)	0	0
Concurrent						
Triangulation	38 (12.9%)	6 (2.0%)	14 (4.8%)	4 (1.4%)	0	0
Embedded	2 (0.7%)	2 (0.7%)	0	0	0	0
Total	184 (62.6%)	32 (10.9%)	60 (20.4%)	18 (6.1%)	0	0

4.5. Sentence structure distribution in RQs across MMR designs

A Chi-square test indicated significant differences in sentence structure distribution across the MMR designs, $\chi^2 (6, N=518) = 86.81, p < .001$, with a medium effect size (Cramér's $V = 0.289$). As shown in Table 5, simple sentence structures dominate (66.0%, $n = 342$), particularly in explanatory designs (49.1%, $n = 168, M = 4.7$), followed by exploratory (32.1%, $n = 110, M = 3.4$) and triangulation (16.9%, $n = 58, M = 2.9$). Complex sentences (30.9%, $n = 160$) are less common, with triangulation and explanatory designs each contributing 28.7% ($n = 46$). Compound sentences are rare (3.1%, $n = 16$), with triangulation designs leading (62.5%, $n = 10$).

As shown in Table 5, simple sentence structures dominate (66.0%, $n = 342$), followed by complex sentences (30.9%, $n = 160$) and compound sentences (3.1%, $n = 16$). Notably, compound-complex sentence structures were entirely absent from the corpus, suggesting a preference among doctoral researchers for less syntactically layered interrogative forms when formulating RQs.

The preference for simple sentences reflects TEFL commitment to clarity, ensuring RQs are accessible to practitioner audiences, such as educators and policymakers. The low use of compound sentences may indicate a deliberate avoidance of syntactic complexity, prioritizing straightforward communication in Iran's EFL context.

Table 5. Sentence structure distribution in RQs across MMR designs

MMR design	Simple sentences (%)	Mean simple	Complex sentences (%)	Mean complex	Compound sentences (%)	Mean compound
sequential						
Explanatory	168 (49.1)	4.7	46 (28.7)	1.3	4 (25.0)	0.11
Exploratory	110 (32.1)	3.4	36 (22.5)	1.1	2 (12.5)	0.05
Concurrent						
Triangulation	58 (16.9)	2.9	46 (28.7)	2.3	10 (62.5)	0.5
Embedded	6 (1.8)	1.5	32 (20.0)	8.0	0 (0.0)	0.0
Total	342 (100.0)		160 (100.0)		16 (100.0)	

Note: Simple Sentences = Total Number of Simple Sentences; Mean Simple = Mean Number of Simple Sentences; Complex Sentences = Total Number of Complex Sentences; Mean Complex = Mean Number of Complex Sentences; Compound Sentences = Total Number of Compound Sentences; Mean Compound = Mean Number of Compound Sentences.

4.6. Verb tense distribution in RQs across MMR designs

A Chi-square test revealed significant differences in verb tense distribution, $\chi^2 (3, N = 518) = 14.28, p < 0.003$, with a small effect size (Cramér's $V = 0.166$). As shown in Table 6, present tense dominates

(99.2%, $n = 514$), with explanatory designs leading (42.4%, $n = 218$, $M = 6.06$), followed by exploratory (28.7%, $n = 148$, $M = 4.63$) and triangulation (21.4%, $n = 110$, $M = 5.50$). Future tense is rare (0.8%, $n = 4$), appearing only in triangulation designs ($M = 0.20$).

The overwhelming use of present tense underscores TEFL focus on ongoing, relevant inquiries, aligning with the practitioner-oriented goals of the field. The rarity of future tense in TEFL may reflect a cultural preference in Iran for addressing immediate pedagogical needs rather than speculative outcomes.

Table 6. Verb tense distribution in RQs across MMR designs

MMR design	Present tense (%)	Mean present tense	Future tense (%)	Mean future tense
sequential				
Explanatory	218 (42.4)	6.06	0 (0.0)	0.00
Exploratory	148 (28.7)	4.63	0 (0.0)	0.00
Concurrent				
Triangulation	110 (21.4)	5.50	4 (100.0)	0.20
Embedded	38 (7.4)	9.50	0 (0.0)	0.00
Total	514 (100.0)		4 (100.0)	

Note: Present Tense = Total Number of Present Tense RQs; Mean Present Tense = Mean Number of Present Tense RQs; Future Tense = Total Number of Future Tense RQs; Mean Future Tense = Mean Number of Future Tense RQs.

4.7. Visual representation of RQ linguistic features and MMR designs

Figure 1 illustrates the prevalence of non-polar Wh-questions, simple sentence structures, and present tense usage in research questions across sequential and concurrent MMR designs, highlighting TEFL preference for clear, descriptive inquiries tailored to practitioner audiences.

Syntactic Features of RQs Across MMR Designs

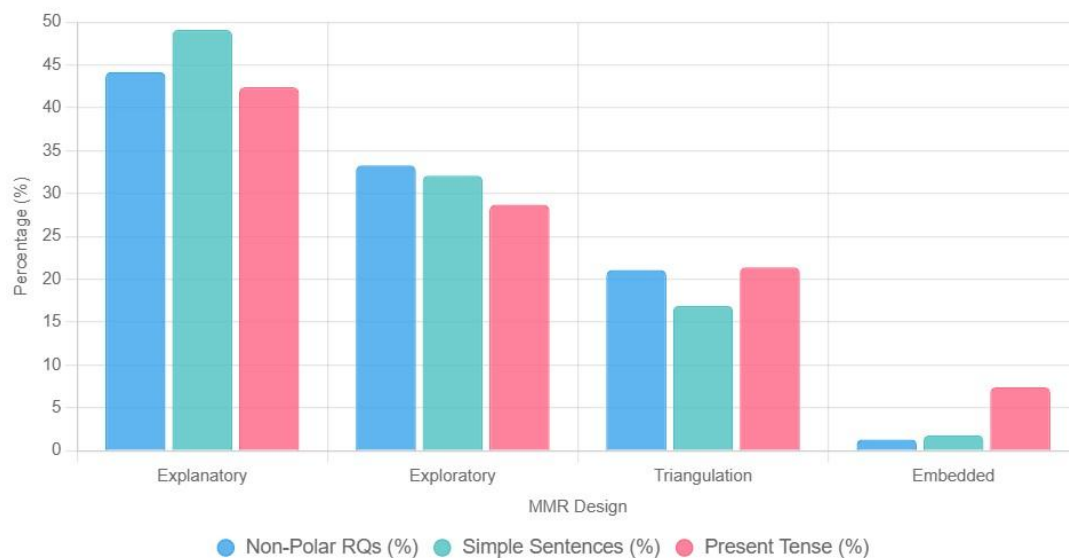


Figure 1. Distribution of key syntactic features of research questions across mixed methods research designs in Iranian TEFL dissertations

V. DISCUSSION

This study provides a comprehensive analysis of mixed methods research (MMR) applications in Teaching English as a Foreign Language (TEFL) PhD dissertations. It reveals methodological and linguistic patterns that illuminate the research practices in the field and evolving identity. By integrating quantitative corpus analysis of 518 research questions (RQs) with qualitative content analysis, the findings offer novel insights into how syntactic features, such as polarity, sentence structure, and verb tense, align with MMR designs, bridging linguistic analysis with methodological integration in TEFL. The preference for sequential MMR designs, with explanatory (39.1%) and exploratory (34.8%) designs comprising nearly three-quarters of the

corpus, aligns with prior observations in educational research (Amini Farsani & Mohammadi, 2022). This predominance reflects the emphasis of TEFL on phased investigations, where quantitative data often provide a foundation for qualitative exploration (e.g., test scores followed by interviews) or qualitative insights inform instrument development, as noted by Creswell and Plano Clark (2018). Unlike psychology, where concurrent designs are common for experimental validation (Venkatesh et al., 2013), the sequential dominance of TEFL may stem from the need to capture cultural and institutional nuances in Iran's EFL context, shaped by national educational policies emphasizing context-sensitive pedagogy (Ebadi et al., 2019).

The linguistic analysis reveals that non-polar Wh-questions dominate (56.8%), particularly factual (62.6%) and evaluative (20.4%) types, reflecting TEFL exploratory and descriptive orientation. Factual questions establish foundational knowledge, while evaluative questions assess pedagogical outcomes, resonating with TEFL practitioner focus (Lim, 2014). The absence of causative and personal Wh-questions suggests a gap in TEFL research, prioritizing descriptive inquiries over causal or learner-specific investigations. This is in contrast to education MMR studies, where causative questions are more common to evaluate intervention impacts (Dawadi et al., 2021). Syntactic features further distinguish TEFL RQs, with simple sentence structures (66.0%) and present tense usage (99.2%) reflecting a commitment to clarity, accessibility, and alignment with standardized academic conventions (Gibson, 2017).

The qualitative content analysis further illuminated the rhetorical work performed by these syntactic choices, revealing three critical functions. The first is Rhetorical Anchoring. The consistent use of simple, present-tense structures is as a linguistic "anchor" that helps novice researchers manage the cognitive load of navigating two distinct paradigms, maintaining coherence amidst methodological complexity. The second is Transparency and Signaling. Specific RQ wording functions as a signaling mechanism; writers use precise syntactic frames to inform examiners which phase, quantitative or qualitative, is currently being addressed, thereby enhancing procedural transparency. The third is Methodological Rationale. The qualitative findings revealed that doctoral writers often employ "evaluative" RQs as a transition tool, providing a necessary bridge to justify how numerical findings necessitate subsequent interview-based interpretations. Together, these functions demonstrate that RQs serve not merely as inquiry statements, but as vital organizers that stabilize the multi-phase nature of mixed-methods designs.

Despite these coherent patterns, persistent challenges in MMR implementation emerge, echoing Riazi and Farsani's (2024) finding that 70.4% of applied linguistics studies lack clear MMR design specifications. Many dissertations failed to articulate how qualitative and quantitative components were integrated, a critical aspect of MMR rigor (Tashakkori & Creswell, 2007). This issue may stem from limited methodological training in Iranian graduate programs or institutional norms prioritizing practical outcomes over methodological clarity. Similarly, only 11% of studies in Merchant et al.'s (2021) review of instructional design MMR properly framed RQs according to MMR principles, indicating a broader challenge in formulating integrative inquiries. These challenges highlight the need for enhanced training in RQ formulation and MMR design, particularly in non-native English contexts like Iran, where linguistic and methodological complexities intersect.

The findings contribute to current scholarship by demonstrating how RQ wording can reveal doctoral researchers' conceptualization of mixed methods in contexts with uneven access to methodological training. At the same time, the limited use of causative and personal Wh-questions highlights opportunities to broaden the functional range of RQs, enabling dissertation writers to address causal, experiential, and learner-centered dimensions of TEFL inquiry. More explicit instruction in design justification, integration strategies, and corpus-informed RQ formulation may support the development of more coherent and theoretically grounded MMR designs.

In sum, while the current linguistic and methodological patterns reflect the disciplinary norms of applied linguistics, persistent challenges in design articulation and integration underscore the importance of targeted methodological support. Enhancing such support can strengthen the quality and transparency of MMR reporting in TEFL doctoral research.



VI. CONCLUSION

This study employing mixed methods research (MMR) to examine how research questions (RQs) are linguistically formulated in 92 Iranian TEFL PhD dissertations. It showed that RQ syntax reflects methodological reasoning and rhetorical convention rather than directly determining methodological integration. Sequential explanatory and exploratory designs were most common, shaped by local dissertation-writing practices. The RQs were predominantly non-polar factual and evaluative Wh-questions, realized in simple sentences and present tense, supporting clarity, accessibility, and coherence across multi-phase MMR designs. The qualitative analysis further highlighted their rhetorical roles in anchoring phases, signaling methodological intent, and enhancing transparency.

The study contributes a replicable framework for analyzing RQ syntax, offers insights into how linguistic formulation reflects doctoral researchers' understanding of MMR, and provides a context-specific account of RQ practices in Iranian TEFL programs. These findings are presented as discipline-specific insights rather than generalizable claims.

Pedagogically, the study underscores the value of explicit instruction in RQ construction within EAP and doctoral writing courses. Corpus-based exemplars, such as those summarized in Table 7, can be incorporated into supervision and methods of training to provide concrete, evidence-based models of effective RQ design, illustrating common syntactic patterns, question types, and rhetorical strategies. Encouraging doctoral researchers to articulate clear rationales for design choices and integration further strengthens methodological transparency and coherence. Table 7, thus, serves as a practical teaching tool, linking analytical findings to applied instructional practice.

While limited to a single national context and surface-level syntactic features, the study highlights some directions for future research, including cross-national comparisons, longitudinal tracking of doctoral RQ development, and experimental evaluation of explicit instructional interventions. Overall, it provides actionable insights for EAP instructors, supervisors, and doctoral researchers aiming to enhance the clarity, coherence, and methodological rigor of RQ formulation in applied linguistics dissertations.

Table 7. Guidelines for formulating research questions in TEFL MMR studies

Aspect	Guideline	Example
Question type	Use non-polar Wh-questions (factual, evaluative) for exploratory depth.	"What are the effects of task-based learning on vocabulary acquisition?"
Sentence structure	Prioritize simple sentences for clarity and accessibility.	"How do teachers perceive flipped classroom strategies?"
Verb tense	Use present tense to emphasize ongoing relevance.	"What factors influence learners' motivation in EFL settings?"
Methodological alignment	Align RQs with MMR design (e.g., factual for exploratory, evaluative for explanatory).	"To what extent does peer feedback improve writing skills?" (explanatory)

In conclusion, this study demonstrates that the syntactic formulation of RQs in Iranian TEFL dissertations is deeply embedded in both local pedagogical norms and broader academic conventions. While the findings provide a clear snapshot of the current practices, they also highlight a critical need for enhanced EAP support. By moving beyond a 'one-size-fits-all' approach to RQ formulation, doctoral researchers can better leverage the linguistic potential of their RQs to signal methodological rigor. Future research should prioritize longitudinal studies to observe how RQ formulation evolves from the proposal stage to the final dissertation defense, providing a more dynamic view of the development of methodological expertise in applied linguistics.

Conflict of interests

The authors had no conflict of interests regarding this study.

AUTHORS' BIOGRAPHIES

Fatemeh Zamanian is a Ph.D. candidate. Her research centers on academic discourse, corpus linguistics, and English for Academic Purposes (EAP). Her current collaborative work with Dr. Karimnia employs mixed-methods corpus analysis to investigate the linguistic and syntactic structures of research.

Amin Karimnia is an Associate Professor of Applied Linguistics at Islamic Azad University, Fa.C., Iran. His research spans English language education, teacher education, translation studies, cross-cultural communication, and language policy. He has published extensively in international journals like *Reading Psychology*, *Semiotica*, and the *Journal of Multicultural Discourses*, and is the author of several books, most recently a 2026 volume on teacher education in Iran published by Palgrave Macmillan.

Zohreh Seifoori is an Associate Professor of ELT at Tehran Islamic Azad University, Science and Research Branch. A Research Board member, and an internationally licensed teacher trainer. She has published a number of research papers in renowned national and international journals. Her research interests include Teacher Education, Learner Autonomy, and Teaching Methodology.

REFERENCES

- Altman, D. G. (1991). *Practical statistics for medical research*. London: Chapman and Hall.
- Amini Farsani, M., & Mohammadi, V. (2022). Mixed methods research in an EFL context: A quality assessment perspective. *Journal of Modern Research in English Language Studies*, 9(2), 99-122.
- Azar, B. S., & Hagen, S. A. (2009). *Understanding and using English grammar*. Pearson Longman.
- Chen, T. Y., & Kuo, C. H. (2012). A genre-based analysis of the information structure of master's theses in applied linguistics. *The Asian ESP Journal*, 8(1), 24-52.
- Collins, K. M., Onwuegbuzie, A. J., & Sutton, I. L. (2006). A model incorporating the rationale and purpose for conducting mixed-methods research in special education and beyond. *Learning Disabilities: A Contemporary Journal*, 4(1), 67-100.
- Creswell, J. W., & Plano Clark, V. L. P. (2018). *Designing and conducting mixed methods research*. SAGE publications.
- Creswell, J. W. (2009). *Research design*. SAGE Publications.
- Dawadi, S., Shrestha, S., & Giri, R. A. (2021). Mixed-methods research: A discussion on its types, challenges, and criticisms. *Journal of Practical Studies in Education*, 2(2), 25-36. <https://doi.org/10.46809/jpse.v2i2.20>
- Dryer, M. S. (2013). Polar questions. In M. S. Dryer, & M. Haspelmath (Eds.), *The World Atlas of Language Structures Online*. Max Plank Institute for Evolutionary Anthropology. <https://doi.org/10.5281/zenodo.13950591>
- Ebadi, S., Salman, A. R., Nguyen, T. T. L., & Weisi, H. (2019). Rhetorical structure variations in abstracts and introductions of applied linguistics master's theses by Iraqi and international students. *Iranian Journal of Language Teaching Research*, 7(1), 101-117.
- Feak, C. B., & Swales, J. M. (2011). *Creating contexts: Writing introductions across genres*. University of Michigan Press.
- Gravetter, F. J., & Wallnau, L. B. (2004). *Statistics for the behavioral sciences: A short course and student manual*. University Press of America.



- Ghanbar, H., & Rezvani, R. (2023). Research questions in applied linguistics research: A microscopic analysis of their distributional and syntactical aspects. *Journal of Research in Applied Linguistics*, 14(1), 156-167. <https://doi.org/10.22055/RALS.2023.18074>
- Gibson, C. B. (2017). Elaboration, generalization, triangulation, and interpretation: On enhancing the value of mixed method research. *Organizational Research Methods*, 20(2), 193-223.
- Hashemi, M. R., & Babaii, E. (2013). Mixed methods research: Toward new research designs in applied linguistics. *The Modern Language Journal*, 97(4), 828-852.
- Hogue, A. (2003). *The essentials of English: A writer's handbook*. New York: Pearson Education.
- 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.
- Kawase, T. (2018). Rhetorical structure of the introductions of applied linguistics PhD theses. *Journal of English for Academic Purposes*, 31, 18-27.
- Kross, J., & Giust, A. (2019). Elements of research questions in relation to qualitative inquiry. *The Qualitative Report*, 24(1), 24-30. <https://nsuworks.nova.edu/tqr/vol24/iss1/2>
- Lim, J. M. (2014). Formulating research questions in experimental doctoral dissertations on Applied Linguistics. *English for Specific Purposes*, 35, 66-88.
- Lim, J. M. H., & Luo, X. (2020). Writing research questions and hypotheses: A genre-based investigation into writers' linguistic resources in social sciences. *ESP Today*, 8(2), 206-226.
- Lu, X., Casal, J. E., & Liu, Y. (2020). The rhetorical functions of syntactically complex sentences in social science research article introductions. *Journal of English for Academic Purposes*, 44, 1-15.
- Lu, X., Yoon, J., & Kisselev, O. (2021). Matching phrase-frames to rhetorical moves in social science research article introductions. *Journal of English for Academic Purposes*, 61, 63-83.
- Merchant, Z. H., Sadaf, A., Olesova, L., & Wu, T. (2021). A systematic review of research questions in mixed methods studies in instructional design. *Pedagogical Research*, 6(4), 1-13, em0107. <https://doi.org/10.29333/pr/11282>
- Onwuegbuzie, A. J., & Leech, N. L. (2006). Linking research questions to mixed methods data analysis procedures1. *The Qualitative Report*, 11(3), 474-498.
- Ozturk, I. (2007). The textual organization of research article introductions in applied linguistics: Variability within a single discipline. *English for Specific Purposes*, 26, 25-38.
- Plano Clark, V. P., & Badiie, M. (2010). *Research questions in mixed methods research*. SAGE handbook of mixed methods in social and behavioral research, 275-338.
- Radford, A. (2009). *An introduction to English sentence structure*. Cambridge University Press.
- Ren, H., & Li, Y. (2011). A comparison study on the rhetorical moves of abstracts in published research articles and master's foreign-language theses. *English Language Teaching*, 4(1), 162-166.
- Riazi, A. M. (2017). *Mixed methods research in language teaching and learning*. Equinox Publishing.

Riazi, A. M., & Candlin, C. N. (2014). Mixed-methods research in language teaching and learning: Opportunities, issues, and challenges. *Language Teaching*, 45, 135-173.

Riazi, A. M., & Farsani, M. A. (2024). Mixed-methods research in applied linguistics: Charting the progress through the second decade of the twenty-first century. *Language Teaching*, 57(2), 143–182. <https://doi.org/10.1017/S0261444823000332>

Shehzad, W. (2011). Outlining purposes, stating the nature of the present research, and listing research questions or hypotheses in academic papers. *Journal of Technical Writing and Communication*, 41, 139-160.

Sheldon, E. (2011). Rhetorical differences in RA introductions written by English L1 and L2 and Castilian Spanish L1 writers. *Journal of English for Academic Purposes*, 10, 238-251.

Soodmand Afshar, H., & Ranjbar, A. (2017). A comparative study of rhetorical moves adopted in research questions and method sections of applied linguistics research articles. *Journal of Research in Applied Linguistics*, 8(1), 46-71.

Soodmand Afshar, H., & Hafez, F. (2021). A mixed-methods investigation of TEFL graduate students' perspectives of qualitative research: challenges and solutions in the spotlight. *The Qualitative Report*, 26(5), 1444-1475. <https://doi.org/10.46743/2160-3715/2021.4614>

Soler-Monreal, C., Carbonell-Olivares, M., & Gil-Salom, L. (2011). A contrastive study of the rhetorical organization of English and Spanish PhD thesis introductions. *English for Specific Purposes*, 30, 4-17.

Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge University Press.

Swales, J. M. (2004). *Research genres: Exploration and applications*. Cambridge University Press.

Tashakkori, A., & Creswell, J. W. (2007). Exploring the nature of research questions in mixed methods research. Editorial. *Journal of Mixed Methods Research*, 1(3), 207-211.

Tashakkori, A., & Teddlie, C. (2003). *The past and future of mixed methods research: From data triangulation to mixed model designs*. In A. Taskakkori, & C.Teddlie (Eds.), *Handbook of Mixed Methods in Social and Behavioral Research* (pp.671-701). SAGE Publications.

Venkatesh, V., Brown, S. A., & Bala, H. (2013). Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems. *MIS Quarterly* 37(1), 21-54.



APENDICES 1

Types of questions and representative RQs from the corpus

135

Question Type	Sample RQ
Polar RQ	Does reflective language learning have any impact on Iranian EFL learners' reflection?(DI ,9)
Non-polar RQ	To what extent do the English textbooks of Iranian senior high school (Vision Series) meet the "cognitive domain" of Bloom's Revised Taxonomy? (DI,34)
Simple Sentence Structure RQ	What is the effect of web-based DA on the grammatical accuracy of high school EFL learners?(DI,22)
Compound Sentence Structure RQ	What is the nature of the variables involved and how do they interact? (DI,3)
Complex Sentence Structure RQ	Is there any significant relationship between digital literacy and the utilization of educational strategies among Iranian English as EFL learners, particularly in terms of their reading comprehension scores? (DI,18)
Simple present tense RQ	What are the learners' perceptions toward using tests to improve their writing performance?(DI,2)
Simple future tense RQ	Will the dynamic WCF produce greater linguistic accuracy, fluency, and complexity, on L2 learners 'use of 'passive voice 'structures when compared to the traditional instructional method? (DI,41)