education sciences

MDPI

Article

Getting Lost in Translation: Examining the Role of Translation with Multilingual Learners

Eleni Pappamihiel [1], Nirmal Ghimire [1,*] and Traci Couts Bellas [2]

[1] Watson College of Education, The University of North Carolina Wilmington, Wilmington, NC 28403, USA; pappamihieln@uncw.edu
[2] Chatham County Schools, Pittsboro, NC 27312, USA; tbellas@chatham.k12.nc.us
* Correspondence: ghimiren@uncw.edu

Abstract

Translation technology has become ubiquitous in multilingual classrooms without evidence-based implementation guidance. This mixed-methods study examined K-12 teachers' translation practices (n = 88 survey; n = 3 district leader interviews), comparing ESL specialists and content teachers to synthesize principles for effective use. Translation use was widespread (81.8%) despite minimal guidance (88.6% lack policies). Common methods included translation applications (89.6%), peer translation (72.2%), and native language texts. ESL specialists reported higher confidence (M = 3.69 vs. 3.18, d = 0.61) and perceived effectiveness (M = 3.76 vs. 3.29, d = 0.56) than content teachers—differences probably attributable to second language acquisition training. Thematic analysis of leader interviews, validated through Structural Topic Modeling, revealed professional development gaps as the strongest convergence (75% alignment). A critical divergence emerged: content teachers rated translation moderately effective, while leaders observed counterproductive practices (11.6% of segments), creating dependency rather than supporting English development. Leaders distinguished productive translation (temporary scaffolding toward English independence) from problematic practices (wholesale content translation). Findings grounded in Contrastive Analysis and Common Underlying Proficiency theory yielded seven evidence-based principles addressing temporary scaffolding, L1 literacy verification, communication versus content contexts, and sustained professional development. The scaffold-versus-crutch framework contributes conceptual clarity for distinguishing productive from counterproductive translation in technology-enhanced multilingual education.

Keywords: machine-based translation; scaffold-versus-crutch framework; translation overuse; second language acquisition; multilingual learner equity; structural topic modeling; mixed method research

Check for updates

Academic Editors: Francisco Perlas Dumanig and Junifer Abatayo

Received: 14 December 2025
Revised: 1 February 2026
Accepted: 4 February 2026
Published: 7 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

A North Carolina content teacher once described himself as heading "Team Translate Everything", believing wholesale translation helped multilingual learners (MLs; also referred to as English learners or ELs in policy contexts) access content. Only after pursuing ESL certification did he realize that strategic, targeted translation proved more effective than translating everything, and that his approach had inadvertently hindered English language development rather than advancing it. His story illustrates the tension at the heart of this study: the same tool can scaffold or undermine language development depending on implementation.

Translation in language education has come full circle. While the Grammar Translation method dominated early language instruction (Hernandez Jaramillo, 2019), the field shifted decisively toward communicative approaches in the 1980s, with translation largely abandoned in favor of methods emphasizing authentic interaction and meaning-making (Richards & Rodgers, 2001). Under the Grammar Translation method, classroom instruction centered on explicit grammar rules, vocabulary memorization, and written translation of literary texts, with little emphasis on oral communication or authentic language use. The communicative turn rejected this approach, prioritizing meaningful interaction, negotiation of meaning, and functional language competence over grammatical accuracy (Savignon, 2017). Today's language classrooms aim to develop both communicative competence and content knowledge, creating tension around translation's role: while wholesale translation contradicts communicative principles, strategic L1 use may support comprehension and participation without replacing target language engagement.

However, artificial intelligence has transformed this landscape. Neural machine translation technologies, far more accurate than early tools, have reintroduced translation into mainstream practice, particularly among content area teachers seeking to provide multilingual learner access to complex academic material. Translation has become ubiquitous in contemporary classrooms, yet this widespread adoption has outpaced the development of evidence-based implementation guidance.

The COVID-19 pandemic accelerated technology integration in language education, with translation tools becoming increasingly accessible and sophisticated (Herrero & Spence, 2023). Post-pandemic, translation technology has become a permanent classroom fixture, with tools like Google Translate integrated into daily instruction. This integration represents a paradigm shift: technology, once supplementary to instruction, now forms the core of delivering lessons and engaging students (Tiwari et al., 2021). As classrooms become more linguistically diverse and technology-rich, educators face a critical question: How can translation support MLs without undermining English language acquisition?

The critical question is not whether to use translation, but how? Under what conditions does translation function as a temporary scaffold supporting language development versus a permanent crutch creating dependency? Despite this question's importance, research on translation in K-12 contexts remains limited, with many studies focusing on adult language learners who bring well-developed native language literacy to the task. We lack systematic evidence about how K-12 teachers actually use translation tools in both ESL and mainstream content classrooms, how practices differ between ESL specialists and content teachers, and what distinguishes productive from counterproductive implementation.

This mixed-method study addresses this gap by examining translation practices among K-12 teachers to develop an evidence-based scaffold-versus-crutch framework. Through survey data (n = 88), district leader interviews (n = 3), and computational validation of expert perspectives, we characterize current implementation patterns, identify the perception gap between teacher beliefs and expert observations, and synthesize seven principles for effective practice that maximize benefits while minimizing dependency. Based on this history and anecdotal accounts, this study attempted to answer the following research questions:

- RQ1: What are the current patterns of translation use among K-12 teachers serving MLs, including tools, modalities, and target student populations?
- RQ2: How do translation practices, confidence, and perceived effectiveness differ between ESL specialists and content area teachers?
- RQ3: How do district ESL leaders distinguish between productive and counterproductive translation practices in K-12 classrooms?

https://doi.org/10.3390/educsci16020263

- RQ4: What evidence-based principles emerge for guiding effective translation use with MLs in public school contexts?

2. Review of the Literature

The concept of translation is not new. Early attempts at machine-based translation were inaccurate and not feasible when compared with human-generated translations. The later shift to neural machine translation (NMT) in the mid-2010s is considered the beginning of truly effective translation.

2.1. AI-Assisted Translation

In the US, scaffolds and accommodations are considered the Gold Standard of instruction for MLs, and for decades now, preservice and in-service teachers have been encouraged to use appropriate scaffolds and accommodations that take language proficiency into consideration. Teachers and schools have attempted to use translation as a scaffold ever since the introduction of Google Translate in 2006 (Turovsky, 2016). Unfortunately, this early version of machine-generated translation was often inaccurate. It was not until 2016 that Google pivoted to deep learning techniques, which significantly improved the accuracy of translations, but even then, the system was only useful for a limited set of languages, such as English, French, Spanish, German, and Chinese. Since then, advances in AI modeling have led to far more accurate translations that are increasingly being used in schools with second-language learners. In most instances, the use of translation was set aside in favor of the previously mentioned communicative methods. Yet, as translation becomes increasingly accessible and more accurate, the question remains: In classrooms where the objective is learning English and content, how is translation being used as a scaffold to learn English rather than a shortcut to understanding content?

2.2. Benefits of Translation

Ducar and Schocket (2018) refer to machine- based translation as the key pedagogical issue of our time. Yet, there remains a great divide between how teachers use this tool, if they use it at all. Many benefits of translation have been found in the research, and translation is viewed positively by both teachers and students (Cook, 2010). From a theoretical standpoint, the use of students' native languages in learning another language is strongly supported. Vygotsky's sociocultural theory posits that learners' native languages are crucial in spanning the Zone of Proximal development (Lantolf et al., 2017). Cummins (1991) has long advocated for language interdependence, and his Common Underlying Proficiency theory supports that learning a new language can be optimized depending on students' ability to successfully transfer skills from their native language to the new language. Increasingly, researchers are discussing the use of translanguaging and exploring how it differs from traditional translation. However, at the heart of both practices is the strategic use of learners' native languages.

Ducar and Schocket (2018) note that Google Translate and the use of translation in general have often been vilified in the classroom, often referred to as a crutch or a cheat. This discouragement of translation was especially strong in the 1980s and 90s when researchers advocated for the monolingual principle and its centrality to communicative competence (Ellis, 1985). However, other literature has supported the strategic use of translation in the acquisition of English in the content areas. These content-area recommendations include using translation to support comprehension, maintain access to grade-level content, build confidence and participation, and facilitate bilingual development, especially in older MLs.

Specific examples of strategic translation include pre-teaching vocabulary with L1 equivalents before reading tasks, providing bilingual glossaries for technical terms,

and allowing brief L1 clarification during collaborative activities before requiring English output (Cook, 2010). Research indicates these practices primarily benefit vocabulary acquisition and reading comprehension rather than grammatical development. Calis and Dikilitas (2012) found that translation tasks improved learners' vocabulary retention and awareness of form-meaning connections, while Mavrou (2020) demonstrated that strategic L1 use reduced cognitive load, freeing working memory resources for higher-order processing of content.

Numerous studies have demonstrated that the structured, strategic use of translation, as in the examples above, can have a positive impact on second language acquisition (Cook, 2010; Howard et al., 2007; Mavrou, 2020; Calis & Dikilitas, 2012; Tati et al., 2024).

Wolf et al. (2008) support the use of translations in grading rubrics and instructions that allow students to better understand expectations and, when appropriate, respond partially in their native language. Such use of translation allows instructors to distinguish content knowledge from their language (English) proficiency. Garcia et al. (2017) advocate for the use of students' home language in supporting classroom discussions to improve understanding of content topics, a practice the authors refer to as translanguaging as opposed to translation.

Several authors advocate for the use of translation and translanguaging to improve content understanding, especially with regard to complex texts. Garcia and Kleyn (2016) and Echevarria et al. (2017) wrote that translation and translanguaging can allow students to access text at a more complex level than in English-only versions. Additionally, these strategies aid in developing comprehension of content-specific vocabulary. However, it is important to note the significant theoretical differences between translation and translanguaging. Translation focuses on moving from one language to another, whereas translanguaging seeks to use linguistic resources across languages simultaneously.

In his 2024 opinion article, Ferlazzo stated that the following considerations must be addressed regarding the specific use of Google Translate.

- Does the use of translation allow students opportunities to express their content knowledge, or does it hinder their development of important content vocabulary?
- Is the translation encouraging confidence in the student or inhibiting engagement?
- Is the translation enhancing the students' understanding?
- Is the translation reducing cognitive load in ways that allow students the mental 'space' to build knowledge?
- Does the use of translation encourage a growth mindset, or does it actually reduce risk-taking by creating overly safe spaces?
- Has translation become the only tool in a teacher's toolbox?

An important point made by Ferlazzo is that teachers and students must always believe that their brains are smarter than the translation tools. Translation should not replace or dictate instruction. At the same time, we should acknowledge that translation offers many potential benefits in the classroom when used thoughtfully and in support of learning.

2.3. Challenges of Using Translation in the Classroom

There are many challenges to the use of both machine-based translation and translation in general. One growing concern relates to the environmental and energy costs associated with the use of AI, an issue that is increasingly highlighted in discussions about responsible technology use (Patterson et al., 2022).

The most obvious technical challenge is the question of translation quality. Groves and Mundt (2015) found that there were persistent errors in machine translation for classroom contexts, especially in the low-instance languages (e.g., Arabic, Pashto, Vietnamese) that

are becoming more and more common in classrooms around the country. AI translations tend to miss subtle differences in meaning and may easily hallucinate outputs when the accurate translation is not readily accessible. Along with the challenge of the translation quality, an additional challenge involves students' ability to effectively interpret and make use of translated text.

Students who have limited or interrupted education (SLIFE) in their native languages often lack sufficient academic skills in their native language to take full advantage of translated materials (Custodio & O'Loughlin, 2017). This presents a challenge that many teachers may not be aware of, particularly when schools or districts lack information about students' native language literacy skills.

In the past few years, translations were criticized for their lack of agility with academic English or with language subtleties (Snow & Uccelli, 2009), but many of these issues have increasingly been resolved as AI translations have become more sophisticated. Yet, other issues remain. As the quality of translation improves, there is a greater chance that teachers and students over-rely on translation as a modification rather than an accommodation or scaffold (Loewen et al., 2020). As teachers rely more heavily on translation, there is an increased need for professional development focused on its effective use. However, research already shows that most teachers do not get sufficient professional development related to working with MLs (Lucas & Villegas, 2011). Given this existing gap, it is unlikely that teachers are receiving adequate PD opportunities for integrating translation with MLs in pedagogically sound ways.

In addition to these concerns, MLs may face social and emotional challenges when using translation. Using translation may introduce feelings of inadequacy. Additionally, students who do not have access to advanced technology or high-speed internet connections may be put at a disadvantage (Warschauer & Matuchniak, 2010).

As noted earlier, PD related to translation is limited. While there is no reliable evidence that assesses how many districts or states in the country provide guidance on the use of translation in classroom instruction, our study suggests that this gap is a significant concern.

2.4. Theoretical Framework: [Conceptualizing Effective Translation] the Scaffold-Crutch Framework

The literature reviewed above reveals a paradox: translation can both support and undermine language development. As Ferlazzo (2024) observes, the critical question is not whether to use translation but how—yet the field lacks operational clarity for distinguishing productive from counterproductive practices. This section establishes the conceptual framework that guides our investigation: the *scaffold-versus-crutch* distinction.

2.4.1. Scaffolding Theory in Second Language

Translation becomes a crutch when these principles are violated: when it is applied wholesale rather than selectively, when it persists indefinitely rather than fading, or when it allows students to bypass English language development entirely. The challenges identified in Section 2.3—particularly Loewen et al.'s (2020) observation regarding the risk of over-reliance takes on new significance through this lens. Improving translation quality paradoxically increases dependency risk when implementation lacks scaffolding principles.

The scaffold-versus-crutch framework complements rather than contradicts translanguaging scholarship, though important distinctions warrant attention. Translanguaging theory (Garcia & Wei, 2014) positions MLs' full linguistic repertoires as resources for meaning-making, rejecting strict language separation. Our framework shares this asset orientation—strategic L1 use supports learning. However, translanguaging scholarship has focused primarily on organic, student-initiated fluid language practices, whereas our framework addresses a distinct phenomenon: teacher-initiated, technology-mediated trans-

lation that may bypass rather than leverage students' linguistic agency. The dependency risk we identify, e.g., students disengaging from English because they expect machine translation, represents a qualitatively different concern than translanguaging's focus on honoring multilingual identities. AI-based translation tools can inadvertently create passive receipt of translated content rather than active translanguaging, particularly when teachers lack training in principled implementation. Thus, our scaffold-crutch distinction offers operational criteria specifically for evaluating technology-mediated translation practices, extending translanguaging principles into contexts where the tool, not the learner, controls language selection.

2.4.2. Teacher Expertise and Implementation Capacity

The capacity to implement translation as a scaffold rather than a crutch depends critically on the teacher's expertise in language acquisition principles. Research demonstrates substantial preparation gaps between ESL specialists and content area teachers. ESL specialists receive focused training in second language development theory, proficiency progression, and principled scaffolding strategies, enabling them to balance L1 support with English exposure and to design instruction that builds toward independence (Janzen, 2008; Lucas & Villegas, 2013; Samson & Collins, 2012).

In contrast, content area teachers, while disciplinary experts, often lack preparation for supporting language learners (Janzen, 2008). Research indicates that fewer than half of teachers with English learners in their classrooms have received formal training on working with MLs (Samson & Collins, 2012). Without language acquisition knowledge, content teachers may conceptualize translation primarily as a content access tool rather than a language development scaffold, unaware that implementation choices affect developmental trajectories. As NCTE (2020) emphasizes, "content area teachers must understand the unique linguistic needs of their ELLs in order to provide meaningful lessons to support their language growth", yet most lack this foundation.

This expertise gap interacts with the PD deficiencies noted in Section 2.3, i.e., if teachers lack general preparation for multilingual learners (Lucas & Villegas, 2011), they are unlikely to receive guidance specific to the principles of translation use. The result is widespread translation implementation without pedagogical grounding in what distinguishes support from dependency.

2.4.3. Toward Evidence-Based Implementation Principles

Despite theoretical clarity that scaffolds should be strategic, temporary, and independence-oriented, operational definitions remain elusive. What makes translation use "strategic" versus "wholesale" in classroom practice? When does temporary support cross into permanent instructional adjustment? At what proficiency level should translation be available, and in what modalities? Teachers make these decisions daily, whether to translate vocabulary, instructions, or entire texts; whether to permit student-initiated translation; whether to restrict translation to newcomers or extend it broadly, yet lack empirical guidance for navigating these choices.

This study addresses this gap by examining how K-12 teachers currently implement translation and how district leaders with language acquisition expertise distinguish productive from counterproductive practices. By integrating teacher-reported patterns with expert pedagogical judgment, we develop evidence-based principles that operationalize the scaffold-versus-crutch distinction for classroom contexts.

3. Materials and Methods

3.1. Research Design

This study employed a convergent parallel mixed methods design (Creswell & Clark, 2017) to examine translation practices among K-12 teachers serving MLs in North Carolina, the United States. In convergent designs, quantitative and qualitative data are collected concurrently, analyzed independently, and then integrated to provide a comprehensive understanding of the phenomenon under investigation. We selected this design because our research questions required both breadth—characterizing prevalence and patterns of translation use across a sample of teachers—and depth—understanding the pedagogical principles that distinguish productive from counterproductive implementation. Neither quantitative nor qualitative methods alone could adequately address our research aims.

The quantitative part consisted of a descriptive cross-sectional survey administered to K-12 teachers (n = 88). Survey data provided information about translation prevalence, methods, target student population, and teacher confidence and effectiveness ratings. We conducted descriptive analyses to characterize patterns (RQ1) and comparative studies to examine differences between ESL specialists and content areas teachers (RQ2). The qualitative component consisted of semi-structured interviews with district-level ESL coordinators (n = 3) to elicit expert perspectives on the implementation of translation (RQ3). Additionally, we analyzed open-ended survey responses using Structural Topic Modeling (STM, Roberts et al., 2019), a computational text analysis method, to identify emergent themes in teachers' descriptions of their scaffolding practices and translation experiences.

Following independent analysis of quantitative and qualitative parts, we integrated findings through joint displays (Guetterman et al., 2015) that explicitly compared quantitative and qualitative results, identifying areas of convergence, divergence, and expansion. This integration enabled us to synthesize evidence-based principles for effective translation use (RQ4), grounded in both empirical and expert practitioner knowledge.

3.2. Context

North Carolina serves a growing population of multilingual learners, with 178,000 students classified as English Learners (ELs) as of November 2024, representing approximately 10.5% of total enrollment (North Carolina Department of Public Instruction, 2024). The largest concentrations of ELs are in urban districts such as Charlotte-Mecklenburg (n = 32,891) and Wake County (n = 21,784), though ELs are present across all 115 local education agencies. While Spanish-speaking students constitute the majority, North Carolina schools serve students speaking over 200 home languages, creating significant linguistic diversity within and across districts.

3.3. Survey Participants

Survey participants (n = 88) were K-12 teachers serving MLs. Recruitment occurred through several venues. A request to participate was distributed through our university's Professional Development Schools (PDS) system that serves 11 school districts and two charter schools. This PDS system includes over 200 partnership schools and over 2000 teachers. We also launched the survey via several social media platforms, including ESL educator groups on Facebook and professional development association listservs. The survey was administered via Google Forms from March to June 2025. Participation was voluntary and anonymous.

The sample consisted of 75% ESL specialists (n = 66) and 25% content area teachers (n = 22). Teachers served across grade levels: 43.2% elementary (grades K-5), 43.2% middle schools (grades 6–8), and 37.5% high school (grades 9–12), with some teachers serving multiple levels. ESL specialists possess expertise in language acquisition and

multilingual education, making their practices particularly relevant for understanding strategic translation use. This sample size, while modest, is appropriate for convergent mixed methods designs where depth of integration matters more than statistical power for population-level inference (Creswell & Clark, 2017).

3.4. Interview Participants

Three district-level ESL coordinators in North Carolina participated in semi-structured interviews (pseudonyms: Florin, Dibya, and Alex). Participants were purposively sampled based on: (a) district-level leadership roles with oversight of ESL programs, (b) experience supporting both ESL specialists and content area teachers, and (c) willingness to discuss translation practices candidly. Participants represented districts varying in size (from <11,690 to >31,000 students), EL population percentage (from 6.71% to 22.94%), and linguistic diversity (from Spanish-dominant to highly multilingual). All three participants had 15+ years of experience in multilingual education, including classroom teaching and instructional leadership. This purposive sampling strategy prioritized depth of expertise over demographic representativeness, appropriate for understanding expert perspectives on translation implementation quality.

3.5. Survey Instrument

The survey consisted of 13 items developed specifically for this study, as no validated instrument existed for examining translation practices in K-12 multilingual contexts (See Appendix A). Development followed an iterative process grounded in practitioner perspectives.

The first author, a member of a professional Facebook group for multilingual learner researchers and practitioners, posted an exploratory prompt asking whether members were observing translation overuse and whether excessive translation undermined English language development. The post generated over 60 substantive responses from ESL specialists, content teachers, and district leaders, revealing key tensions that informed survey construction. Pseudonyms are used below to protect the respondent's identity.

Practitioners described translation overuse as "extremely common" with "well-intentioned educators wanting to help newly arrived students but not understanding that over-translating does not help to develop language" (Commentor V). Several respondents articulated the scaffold-crutch distinction that became central to our framework: "The new flood of translation devices is creating the perception that you can translate rather than scaffold" (Commentor C). Others questioned L1 literacy assumptions: "Can they even read at that level in Spanish?" (Commentor G). One teacher's reflection captured the perception gap our study later documented: "Last year, I thought giving translations would help my students have access to the lessons. I feel like they barely learned anything. . . I realized that every single translation I did robbed them of an opportunity to learn English" (Commentor J). These practitioner insights, combined with literature review, university discussions regarding translation use, review of district guidelines (New York State Guidance, Brunswick County School, NY), and pilot testing with a few ESL teachers, informed the final survey items.

Survey items addressed: teacher role and grade levels served (Items 1–2); translation use and reasons for non-use (Items 3–4); student L1 use for socialization (Item 5); target proficiency levels and translation methods (Items 6–8, 5-point Likert scale: *1 = not at all*, *5 = extremely*); other instructional scaffolds used (Item 11; open-ended); availability of district/school translation guidelines (Item 12); and additional comments about translation practices (Item 13; open-ended).

3.6. Interview Protocol

Semi-structured interviews followed a protocol designed to elicit district leaders' perspectives on translation implementation across classroom contexts (Appendix B). The social media responses described above also informed protocol development, as practitioners raised questions about differences between ESL and content teacher practices, the absence of district guidance, and whether translation was advancing or impeding English acquisition. These practitioner-identified tensions shaped our interview questions, which addressed: current translation practices and institutional support, observations of translation use by teacher type (ESL versus general education), and leaders' evaluative perspectives on whether translation advances English language acquisition.

The protocol included follow-up probes for questions requiring elaboration or clarification. Questions were sequenced to move from descriptive (current practices) to evaluative (effectiveness judgements), allowing participants to ground their assessments in concrete observations. All interviews were conducted via video conference (Zoom) in February 2025, lasted 13–27 min (M = 20 min), and were audio-recorded with participant consent. Recordings were transcribed verbatim using a professional transcription service, then verified for accuracy by the research team.

3.7. Data Analysis

3.7.1. Quantitative Analysis

Survey data were analyzed using R version 4.5.2 (R Core Team, 2025) with packages including tidyverse (Wickham et al., 2019), effsize (Torchiano, 2020), and psych (Revelle, 2025). Analysis proceeded in two phases.

Phase 1: Descriptive analysis. We calculated frequencies and percentages for categorical variables (translation use, methods, target populations, guideline availability) and means, standard deviations, and medians for continuous variables (confidence, effectiveness ratings). For multi-select items (e.g., translation methods), we created binary variables indicating whether each option was selected and reported percentages based on the relevant subsample (e.g., teachers who use translation). Cross-tabulations examined translation practices by teacher role and grade level. These analyses addressed RQ1.

Phase 2: Comparative analysis. We compared ESL specialists and content area teachers using Welch's t-tests for continuous variables (confidence, effectiveness) and Fisher's exact test for categorical variables (translation methods). Welch's t-test was selected because it does not assume equal variances, appropriate given unequal group sizes (ESL n = 66, General n = 22). We calculated Cohen's d effect sizes for mean differences, interpreting d = 0.2 as small, 0.5 as medium, and 0.8 as large (Cohen, 1988). Given the exploratory nature of this study and modest sample size, we emphasize effect sizes and confidence intervals over p-values, consistent with recommendations for transparent reporting of exploratory research (Cumming, 2014). These analyses addressed RQ2.

Assumptions for t-tests (normality, independence) were assessed through visual inspection of Q-Q plots and consideration of sampling procedures. While confidence and effectiveness ratings showed slight negative skew, t-tests are robust to moderate departures from normality, particularly with samples exceeding 30 per group (Boneau, 1960). Independence was reasonably assumed given anonymous survey responses from different schools/districts.

We initially explored associations between translation practices and grade level taught, as well as guideline availability. However, grade level data included substantial overlap (teachers often serve multiple levels), precluding independent group comparisons. Only 10 teachers (11.4%) reported having district guidelines, insufficient for robust statistical com-

parison. Therefore, guideline-related findings are reported descriptively with appropriate caveats about the small sample size.

3.7.2. Structural Topic Modeling

Open-ended survey responses were analyzed using Structural Topic Modeling (STM; Roberts et al., 2019) implemented via the stm package in R. STM is a probabilistic text analysis method that identifies latent topics—recurring patterns of co-occurring words—within a corpus. Unlike manual coding, STM analyzes all documents simultaneously, providing systematic and replicable topic identification. We applied STM separately to responses from two open-ended items: Item 11, asking about other instructional scaffolds (n = 68 responses after excluding non-informative responses such as "N/A" or "none"), and Item 14, soliciting additional comments about translation (n = 37 responses).

Text preprocessing followed standard procedures. Rare words appearing in fewer than two documents were removed from the analysis. The optimal number of topics (K) was determined using searchK() across a range of values (K = 6–9 for Item 11; K = 5–8 for Item 14), selecting K that maximized semantic coherence and exclusivity. Final models used K = 8 for Item 11 and K = 6 for Item 14, both initialized using spectral initialization and converged via variational Expectation-Maximization (final approximate per-word log-likelihood bounds: Item 11 = −3.226, Item 14 = −3.177).

Topics were labeled using FREX terms, which balance word frequency and exclusivity (Bischof & Airoldi, 2012). Topic prevalence was calculated as the mean proportion (Θ) across documents. For each topic, we extracted exemplary quotations using findThoughts(), which identifies documents with the highest posterior probability for that topic. To assess model robustness, we examined convergence diagnostics, compared alternative K values, and inspected topic correlations. Sensitivity analyses with alternative K specifications yielded substantially similar themes, supporting interpretation stability. STM results provided contextualization for quantitative findings and contributed to our synthesis of evidence-based principles.

3.7.3. Qualitative Analysis

Interview transcripts were analyzed using thematic analysis (Braun & Clarke, 2006). The first author read transcripts iteratively, developing codes that were grouped into candidate themes through discussion with the second author. This process yielded seven themes: STRATEGIC_ESL (pedagogical use of translation and native language), GUIDELINES (district policies and infrastructure needs), EQUITY (fairness and access concerns), MISUSE_CONTENT (inappropriate translation in content classrooms), TECHNOLOGY (tool use and quality), SCAFFOLD_CRUTCH (temporary support versus permanent dependency), and PD_GAP (professional development needs).

To triangulate findings and assess whether manually identified themes reflect distinct linguistic patterns, we segmented transcripts by question-response turns (n = 250 segments) and applied Structural Topic Modeling. This dual-method approach enabled complementary analysis: manual coding captures conceptual themes (what segments are about), while STM identifies linguistic topics (recurring word co-occurrence patterns). We assessed alignment between STM-derived topics and manually coded themes by calculating the proportion of overlap between the two coding systems, with alignment strength interpreted using thresholds of $\geq$70% (strong), 60–69% (moderate), and <60% (weak) convergence. We anticipated partial rather than complete alignment, as conceptually rich themes such as SCAFFOLD_CRUTCH or GUIDELINES may be discussed through examples and implicit references rather than distinct vocabulary clusters. This dual-method approach enabled us to assess whether manually identified themes reflected distinct linguistic patterns (cap-

tured by STM) while recognizing that manual coding may capture conceptual relationships beyond word co-occurrence. This analysis addressed RQ3.

3.7.4. Integration

Following independent analysis of quantitative and qualitative strands, we integrated findings using two strategies. First, we created a joint display (Guetterman et al., 2015)—a table explicitly comparing survey results and interview themes—to identify convergence (where findings from both strands aligned), divergence (where they differed), and expansion (where one strand provided additional insight not captured by the other).

Second, we used qualitative findings to explain quantitative patterns. For instance, ESL teachers reported higher confidence and effectiveness than content teachers, a pattern explained by leaders' observations that ESL specialists receive training in language acquisition principles while content teachers lack this foundation.

Finally, we synthesized across all data sources—survey patterns, STM topics, interview themes, and district guidelines—to derive evidence-based principles for effective translation use (RQ4). This synthesis represents the study's primary contribution, integrating empirical description with expert wisdom to generate actionable guidance for practitioners.

3.8. Ethical Consideration

This study was reviewed and approved by the University of North Carolina Wilmington institutional review board (Application #H25-0767). Survey participants provided informed consent by proceeding past an information page explaining the study purposes, voluntary nature, and data confidentiality. Interview participants provided verbal informed consent, which was audio-recorded at the beginning of each interview. All data were stored on a password-protected Google Drive accessible to the research team. Survey data were anonymous; interview participants were assigned pseudonyms to protect confidentiality. Participants received no compensation for participation.

4. Results

This study examined translation practices among K-12 teachers serving MLs in North Carolina through a convergent parallel mixed methods design. Quantitative survey data (n = 88) revealed widespread translation use despite minimal institutional guidance, with technology-mediated approaches dominating implementation. ESL specialists and content area teachers employed similar methods but differed in confidence and perceived effectiveness, with medium effect sizes favoring ESL teachers.

4.1. Translation Prevalence and Patterns (RQ1)

The sample comprised predominantly ESL specialists (75.0%, n = 66) with content area teachers representing 25.0% (n = 22). Translation use was widespread: 81.8% (n = 72) reported using translation as an instructional scaffold, while only 18.2% (n = 16) did not. Despite high prevalence, institutional support was minimal: 88.6% (n = 78) lacked district guidelines. Most teachers (78.4%) allowed native language socialization. Table 1 presents complete participant characteristics.

Given the predominance of ESL specialists in the sample, findings regarding translation prevalence, methods, and target populations primarily reflect ESL specialist perspectives: content teacher patterns should be interpreted cautiously give the smaller subsample (n = 22). Among non-users, half (50.0%, n = 8) reported that translation interferes with English acquisition, followed by discomfort using unfamiliar language (31.2%, n = 5) and policy restrictions (31.2%, n = 5). Notably, no teachers endorsed English-only ideology, indicating opposition stems from pedagogical concerns rather than linguistic restrictivism.

Table 1. Participant characteristics and translation use patterns (n = 88).

Characteristics	N	%
Teacher Type		
ESL teacher	66	75.0
Content area/General education teacher	22	25.0
Grade Levels Taught		
Elementary	38	43.2
Middle grades	38	43.2
High school	33	37.5
Translation Use		
Yes	72	81.8
No	16	18.2
Reason for not Using Translation		
Uncomfortable using a language I do not know	5	31.2
Ineffective as an instructional strategy	4	25.0
Not allowed to use translation as a scaffold	5	31.2
Interferes with students' English learning	8	50.0
Belief that only English should be used	0	0.0
Allow Native Language Socialization		
Yes	69	78.4
Maybe/Sometimes/Conditional	3	3.4
No	1	1.1
Not applicable	16	18.2
District Guidelines Provided		
Yes	10	11.4
No	78	88.6

Note. Grade level and reasons for not using translation percentages sum to more than 100% because teachers could select multiple items.

Among 72 teachers using translation, implementation was technology-mediated. The most common methods were peer translation (72.2%, n = 52), student use of translation apps (59.7%, n = 43), and teacher use of apps for text translation (58.3%, n = 42). Teachers also provided native language texts (50.0%, n = 36), used apps for spoken instructions (51.4%, n = 37), and accessed dictionaries (29.2%, n = 21). Regarding generation mechanisms, student-generated translations (55.6%, n = 40) and software/apps (72.2%, n = 52) predominated, while teacher-generated translations (47.2%, n = 34) and simultaneous devices (6.9%, n = 5) were less common.

Translation extends across proficiency levels. In North Carolina, English proficiency level is determined by an ML's WIDA score, with Level 1 being an entry level of proficiency and Level 6 defined as the highest level of English proficiency. While 19.4% (n = 14) used translation only with newcomers and 29.2% (n = 21) primarily with newcomers and Level 2 students, substantial proportions used it more broadly: 26.4% (n = 19) allowed student self-selection and 20.8% (n = 15) offered it at any level based on teacher judgment.

Overall confidence in translation accuracy averaged 3.57(SD = 0.87) and perceived effectiveness 3.65 (SD = 0.84) on a 1–5 scale. These moderate ratings suggest general satisfaction but room for improvement through enhanced quality or pedagogical training. Ratings varied minimally across elementary (confidence M = 3.52; effectiveness M = 3.68), middle (3.48; 3.58), and high school (3.67; 3.63) levels.

4.2. Role Differences in Confidence and Effectiveness (RQ2)

Fisher's exact tests comparing ESL specialists (n = 55) and content area teachers (n = 17) who use translation revealed no significant differences across translation meth-

ods, generation approaches, or target proficiency levels (all $ps > 0.097$; Table 2). Both groups showed comparable adoption of student apps (ESL 61.8%, content 52.9%), native language texts (ESL 47.3%, content 58.8%), peer translation (ESL 72.7%, content 70.6%), and other approaches.

Table 2. Translation methods and practices by teacher role.

Practice	ESL Teachers n (%)	Content Area Teachers n (%)	Fisher's Exact *p*
Translation Methods	(*n* = 55)	(*n* = 17)	
I use an app to translate text	31 (56.4%)	11 (64.7%)	0.587
Students use translation apps	34 (61.8%)	9 (52.9%)	0.578
I provide access to texts in the native language	26 (47.3%)	10 (58.8%)	0.580
Students use native language dictionaries	17 (30.9%)	4 (23.5%)	0.762
Students translate for each other	40 (72.7%)	12 (70.6%)	1.00
I use an app to translate spoken instructions	25 (45.5%)	12 (70.6%)	0.097
How Translation is Generated			
Teacher translates (speaking and/or writing)	28 (50.9%)	6 (35.3%)	0.283
Teacher uses software or an app to translate	37 (67.3%)	15 (88.2%)	0.125
Students generate translation as needed	32 (58.2%)	8 (47.1%)	0.578
Students use simultaneous translation devices	5 (9.1%)	0 (0.0%)	0.331
Target Proficiency Levels			
Only with newcomers	11 (28.2%)	3 (20.0%)	0.733
Mostly newcomers, but also Level 2s	16 (41.0%)	5 (33.3%)	0.758
Anyone who feels they need it	14 (35.9%)	5 (33.3%)	1.0
Anyone I feel needs it at any level	9 (23.1%)	6 (40.0%)	0.309

Note. Percentages based on teachers who use translation ($n = 72$). Multiple selection was allowed for all categories, so percentages do not sum to 100%. Fisher's exact test was used due to small expected cell frequencies.

Despite methodological similarities, Welch's *t*-tests identified meaningful differences in confidence and effectiveness. ESL teachers reported higher confidence ($M = 3.69$, $SD = 0.79$) than content teachers ($M = 3.18$, $SD = 1.01$), $t(22.34) = -1.92$, $p = 0.068$, $d = 0.61$, 95% CI [−1.17, −0.05]. Similarly, ESL teachers rated translation as more effective ($M = 3.76$, $SD = 0.79$) than content teachers ($M = 3.29$, $SD = 0.92$), $t(23.82) = -1.90$, $p = 0.069$, $d = 0.56$, 95% CI [−1.13, −0.01]. Both effect sizes fall in the medium range (Cohen, 1988), suggesting potentially substantive differences: ESL teachers report approximately one-half to two-thirds standard deviation higher confidence and effectiveness. Given the small content-teacher subsample ($n = 22$) and marginal *p*-values (0.068, 0.069), these patterns should be interpreted as suggestive rather than confirmatory, indicating directions for future research with larger, more balanced samples. Table 3 presents complete comparisons.

Assumption checks confirmed acceptable conditions for Welch's *t*-tests. Levene's tests were non-significant (confidence: $F(1, 70) = 0.20$, $p = 0.66$; effectiveness: $F(1, 70) = 0.33$, $p = 0.57$), with variance ratios below 2.0. Both variables showed negative skew, but *t*-tests are robust to moderate normality departures when sample sizes exceed 30 (Boneau, 1960). The ESL subsample ($n = 66$) met this criterion; the content subsample ($n = 22$) fell slightly below, acknowledged as a limitation.

Grade level analyses were descriptive only due to substantial overlap—teachers serving multiple bands—precluding independent comparisons. The guidelines subsample ($n = 10$) was insufficient for statistical comparison, though all 10 teachers with guidelines used translation, suggesting guidelines may facilitate implementation.

Table 3. Comparison of confidence and effectiveness ratings between ESL and content area teachers.

Variable	ESL Teachers *M* (*SD*) (*n* = 66)	Content Area Teachers *M* (*SD*) (*n* = 22)	*t* (*df*)	*p*	Cohen's *d* [95% CI]
Confidence in Translation Accuracy					
Overall rating (1–5 scale)	3.69 (0.79)	3.18 (1.01)	−1.92 (22.34)	0.068	0.61 [−1.17, −0.05]
Effectiveness of Translation					
Overall rating (1–5 scale)	3.76 (0.79)	3.29 (0.92)	−1.90 (23.82)	0.069	0.56 [−1.13, −0.01]

Note. Welch's *t*-test was used due to unequal group sizes and variances. Confidence intervals for Cohen's *d* calculated using the noncentral t-distribution. Effect size interpretation: 0.2 = small, 0.5 = medium, 0.8 = large (Cohen, 1988). Despite *p*-values > 0.05, medium effect sizes suggest ESL teachers report mostly higher confidence and perceived effectiveness than content area teachers. Given exploratory aims and unequal group sizes (ESL *n* = 66, content *n* = 22), findings are suggestive and warrant replication with larger content-teacher samples.

4.3. District ESL Leaders' Characterization of Translation Implementation (RQ3)

Semi-structured interviews with three district-level ESL coordinators yielded 250 coded leader response segments. Manual thematic analysis identified seven themes characterizing how leaders distinguish productive from counterproductive translation practices: STRATEGIC_ESL (pedagogical use of translation and native language, *n* = 67, 26.8%), GUIDELINES (district policies and infrastructure needs, *n* = 53, 21.2%), EQUITY (fairness and access concerns, *n* = 34, 13.6%), MISUSE_CONTENT (inappropriate translation in content classrooms, *n* = 29, 11.6%), TECHNOLOGY (tool use and quality, *n* = 26, 10.4%), SCAFFOLD_CRUTCH (temporary support versus permanent dependency, *n* = 25, 10.0%), and PD_GAP (professional development needs, *n* = 16, 6.4%). Table 4 presents the complete distribution.

Table 4. Prevalence of themes in district ESL leaders' interviews.

Theme (STM Topics)	Manual Coding %	Coded Segments (n)	STM Prevalence %
STRATEGIC_ESL (T1, T5, T6)	26.8	67	18.98
GUIDELINES (T10)	21.2	53	3.65
EQUITY (T3)	13.6	34	4.75
MISUSE_CONTENT (T8)	11.6	29	5.01
TECHNOLOGY (T2, T4)	10.4	26	6.29
SCAFFOLD_CRUTCH (T7)	10.0	25	3.46
PD_GAP (T9)	6.4	16	3.49

Note. Manual coding percentages represent the proportion of 250 coded segments assigned to each theme. STM prevalence reflects the proportion of STM-derived topic loadings associated with each theme after topic-theme mapping. STM = Structural Topic Modeling with K = 10 topics.

4.3.1. Computational Validation via Structural Topic Modeling (STM)

To validate manual coding, we applied Structural Topic Modeling (STM) to the same corpus following procedures parallel to survey analysis (as described in Section 3.7.2). The STM identified K = 10 topics (selected via searchK maximizing coherence and exclusivity) that mapped conceptually to the seven manual themes. However, alignment between STM-derived and manually coded themes was modest (49.3% overall), with substantial variation across themes (Table 5). Only PD_GAP showed strong convergence (75.0%), while GUIDELINES (35.8%), TECHNOLOGY (38.5%), and EQUITY (44.1%) showed weak alignment.

Table 5. Alignment between manual coding and STM-derived themes.

Manual Theme	STM Alignment (%)	Primary STM Confusion	Key Observation
STRATEGIC_ESL	49.3%	Self-aligned	Theme spans multiple linguistic patterns
PD_GAP	75.0%	STRATEGIC_ESL (25%)	Clear linguistic signature
MISUSE_CONTENT	58.6%	STRATEGIC_ESL (22%)	Often discussed within ESL context
EQUITY	44.1%	STRATEGIC_ESL (35%)	Implicit theme woven through discussion
SCAFFOLD_CRUTCH	44.0%	STRATEGIC_ESL (44%)	Conceptual framework vs. explicit terms
GUIDELINES	35.8%	STRATEGIC_ESL (36%)	Discussed implicitly across topics
TECHNOLOGY	38.5%	STRATEGIC_ESL (39%)	Technology use embedded in practice discussion
Overall Alignment	49.3%	-	-

Note. Alignment values indicate the percentage of manually coded segments correctly classified by STM-derived theme assignment. Higher alignment indicates stronger agreement between manual interpretation and probabilistic topic modeling. STM = Structural Topic Modeling.

This divergence reveals an important methodological insight: STM captures *linguistic topics* (co-occurring words), while manual coding identifies *conceptual themes* (what segments are about). Themes like GUIDELINES and SCAFFOLD_CRUTCH represent conceptual frameworks leaders apply across discussions rather than distinct vocabularies, explaining their diffuse STM signatures. Conversely, PD_GAP's strong alignment reflects its consistent linguistic markers ("need help", "requests", "support"). Given this divergence, we report manual coding as the primary thematic structure, using STM results to identify sub-themes within STRATEGIC_ESL and to validate that manually identified themes reflect distinct discourse patterns, even if not always captured by word co-occurrence alone. This divergence is methodologically expected rather than problematic. STM serves a complementary rather than confirmatory role: strong alignment (e.g., PD_GAP at 75%) indicates themes with consistent linguistic markers, while weaker alignment (e.g., GUIDELINES at 35.8%) reveals themes that coders identified through conceptual interpretation of implicit content. Both patterns provide valid evidence, and together they offer a richer understanding than either method alone.

4.3.2. Strategic Use of Translation vs. Productive ESL Practice

The most prevalent theme, STRATEGIC_ESL (26.8% manual; 19.0% STM; 49.3% alignment), reflected leaders' consensus that translation is beneficial when used as a temporary scaffold to support meaning-making and language development. STM analysis revealed that this theme encompasses three distinct sub-practices: (a) native language as pedagogical tool (Topic 1, 6.7% prevalence, FREX: language, native language, speak another language), (b) vocabulary and practice activities (Topic 5, 5.9%, FREX: word, practice, activity), and (c) English participation strategies with parent engagement (Topic 6, 6.4%, FREX: English, parents, participate). This topic proliferation validates the richness of the strategic ESL practices leaders described. This theme of strategic competence can be easily seen in the comments below.

Florin noted,

> "Students have an opportunity to see the word, visualize the word, and connect it to their first language... but we never rely fully on translation. The goal is learning English."

Dibya emphasized that translation should assist language analysis, not replace instruction:

> "When the student is using English, our job is to understand where their errors are coming from in that process of language acquisition, not to replace instruction with translation."

She illustrated this principle through her own experience learning Spanish with middle school students:

> "When I was taking Spanish lessons, I told my middle schoolers, and boy, were they hard on me. They would check my homework... But what I did in that process was it was a beautiful building of a relationship with them in putting myself in their shoes of learning English, and saying that, hey, it's okay, we're not perfect, but we're all learning here."

Alex similarly acknowledged selective benefits:

> "Google Translate can be a fantastic tool when you use the oral piece with a partner—it just isn't as expansive as students would like."

Together, these comments show that leaders conceptualize translation as complementary to, rather than a substitute for, English instruction. The moderate STM alignment reflects STRATEGIC_ESL's breadth—spanning native language use, vocabulary development, and participation strategies—rather than representing a single linguistic pattern.

4.3.3. Equity and Linguistic Access

As shown in Table 4, EQUITY accounted for 13.6% of manual codes (4.7% STM; 44.1% alignment). Leaders described how translation practices can unintentionally favor Spanish-speaking students while marginalizing those who speak other languages. The weak STM alignment suggests equity concerns are often discussed implicitly within other themes rather than through distinct terminology. STM Topic 3 (4.7% prevalence, FREX: Spanish, Spanish speaking, Arabic, group) captured only explicit mention of Spanish privilege, while leaders' equity lens permeated broader discussions of practice.

Florin explained:

> "Just because a student speaks Spanish doesn't mean they are literate in it... We're seeing lots of students who have that gap."

She also highlighted mixed-language classrooms:

> "We have very mixed groupings now... so it doesn't work to have a connection with one group of the room and not the others."

Dibya illustrated the mathematics of inequity:

> "By your [sic] using Spanish, because 15 kids are Spanish speakers and eight are not, these eight are getting even less."

She further explained systemic disparities:

> "We have a robust Spanish-speaking interpreter group, but that doesn't meet the needs of families who speak other languages."

Leaders emphasized that equitable translation infrastructures must support all linguistic groups. This theme triangulated with survey findings showing Spanish-dominant methods (72.2% use peer translation; Table 2), confirming that the Spanish privilege concern emerged from observed practice patterns, not merely leader perception.

4.3.4. Technology as Support and Barrier

The TECHNOLOGY theme (10.4% manual; 6.3% STM; 38.5% alignment) revealed a sharp contrast between productive and problematic use of technological translation tools.

STM analysis distinguished two sub-topics: (a) Google Translate copy-paste practices (Topic 2, 2.8%, FREX: copy paste, google translate, translate), and (b) Language Line phone interpretation (Topic 4, 3.5%, FREX: phone, call, speaker, person). The modest alignment (38.5%) reflects that leaders often discussed technology quality within broader pedagogical contexts rather than as a standalone topic.

Problematic uses centered on the copy-and-paste translation of instructional content. Florin described:

> "Kids are savvy—they will copy and paste from Google Translate, and teachers do the same. . . When you depend on it for whole texts, the translation is often wrong, and students end up doing double work."

Dibya emphasized the broken language problem:

> "When you go in, and you listen to somebody who's using totally wrong language and the kids are snickering, the teacher is totally unaware, but no learning is happening."

In contrast, Language Line, which is a pay-as-you-go, person-to-person translation service used for school meetings, was consistently viewed as an equitable communication tool:

> "The interpreter connects both sides for conversation. . . Language line offers over 240 languages. All of our languages are now covered. It opens communication in ways Google Translate cannot."
>
> (Florin)

Alex added nuance, explaining limitations of certain apps, "The database just isn't as expansive as students would like."

Thus, Language Line technology is considered highly productive for family communication because it presents opportunities for real-time negotiation of meaning, but its cost is prohibitive for use in the classroom to translate full texts or assignments. Additionally, the purpose of each context is different—family communication versus English language acquisition. This distinction—communication tools versus content translation—provides a critical framework for evaluating technology appropriateness.

4.3.5. Misuse of Translation for Content Delivery

MISUSE_CONTENT accounted for 11.6% of manual codes (5.0% STM; 58.6% alignment, Table 5). This strong alignment indicates misuse has a consistent linguistic signature (STM Topic 8, FREX: translation, classroom, talk, meet), suggesting leaders discuss counterproductive practices with recognizable patterns. Leaders described widespread translation practices that increased cognitive load and impeded English development.

Florin explained:

> "Teachers will take a reading passage or a social studies article and drop it into Google Translate and say, 'All right. Here's your assignment'. . . Students are doing double the work because they don't read in their home language, and it's not an exact translation."

She further described the cascade of problems:

> "Teachers think, 'I did the most amazing thing. I turned the entire unit into Spanish', and then they flop it on their desk, and the student looks up like a deer in headlights. . . They don't have the words to admit it, but they actually don't know how to read it."

Dibya added:

> "Some teachers believe that if they speak another language, they should be tested in that language—even when the instruction is not in that language."

These examples show how translation becomes counterproductive when it replaces appropriate instructional scaffolding. This theme's strong alignment with STM and its convergence with survey findings—where content teachers reported lower confidence (M = 3.18) despite perceiving moderate effectiveness (M = 3.29)—suggests a perception gap: teachers believe translated materials help, while leaders observe they create barriers.

4.3.6. Need for District Guidelines

As shown in Table 4, GUIDELINES was the second most frequent manual theme (21.2%), though its STM prevalence was low (3.7%, FREX: door, school, district, member, staff) and alignment was modest (35.8%). This substantial divergence—the largest in Table 5—reveals that guidelines are discussed implicitly throughout interviews rather than as a distinct topic. The STM captured only explicit mentions of district infrastructure and family access (front office, door), while leaders' calls for guidelines permeated discussions of equity, misuse, and PD gaps. This pattern explains why a theme comprising 21% of manual codes yielded only 3.7% STM prevalence: leaders frame guideline needs through examples of problems rather than explicit policy discussion. Professional development in this area needs to focus on solutions and strengths-based use of translation and pull teachers away from a reactive use of the tools.

Florin noted, "We had to clamp down on this... A teacher's first reaction is, 'I'll just translate everything.' That's not an expectation." She also described streamlining translation support, "We realized we needed a system... so we created a Google form to determine what support is actually needed."

Dibya described developing scenario-based guidance:

> "We put together guidelines with scenarios, and by also mainly telling them that let's model good language... There was a little more pushback than I expected, because especially in secondary, they feel that something is better than nothing."

Across districts, leaders stressed the need for clear, scenario-based policies that distinguish productive from counterproductive translation practices. This theme triangulates with survey findings showing that only 11.4% of teachers have district guidelines, confirming a widespread infrastructure gap.

4.3.7. Translation as Scaffold Versus Crutch

The SCAFFOLD_CRUTCH theme (10.0% manual; 3.5% STM; 44.0% alignment) offered a conceptual distinction that leaders felt teachers often lacked. The weak STM alignment (44.0%) reflects that the scaffold-crutch is an evaluative framework leaders apply to assess practices rather than a topic with distinct vocabulary. STM Topic 7 (FREX: instruction time, pull, program, interpret) captured the discussion of instructional structure but not the conceptual distinction itself. Translation was viewed as productive when it enabled understanding but counterproductive when it fostered dependency.

Florin described this pattern as "Students will sit back and say, 'I don't have to listen because my teacher will translate everything'. That becomes a crutch, not a scaffold." She further explained the modification approach:

> "I will ask the question, What give me an example of this week. What have you assigned everyone else? And then we go through ways that we can modify them so that the MLs can participate. Teachers are usually shocked that it's okay to modify your grading for this student!

Dibya reinforced that students must engage with English, "Students should be meeting at the bridge—bringing small pieces of English to understand, not relying fully on translation." She described allowing L1 to be used strategically:

> "Our guidance has been allowing them to use their own language as a crutch, if they need to clarify, so that they can get their thoughts out. But as a teacher, to put themselves in a position of a learner."

Alex's comment that translation is needed "from time to time" suggests similar boundaries. This scaffold-crutch framework represents a key conceptual contribution from the qualitative data—a distinction not captured in survey responses but essential for evaluating translation quality.

4.3.8. Professional Development Gaps

Finally, PD_GAP (6.4% manual.; 3.5% STM) showed the highest alignment with STM (75%; Table 5). This strong convergence indicates PD needs have a consistent linguistic signature (Topic 9, FREX: need help, they need, get requests, google form), likely because leaders quote teachers' explicit help requests. Leaders noted that teachers often request translation when what they need is guidance in modifying assignments.

Florin explained, "Teachers say, 'I need translation', when what they really need is help modifying what everyone else is doing." She highlighted the impact of individualized coaching, "It's far superior to talk through these issues during planning. . . that's where we really work through what the classroom needs."

Dibya framed this as a mindset issue: "They'll say, but I don't know this language, so I can't teach them. And I always tell them, y'all, your superpower is teaching them English. That part, you know."

Leaders emphasized PD as essential for helping teachers understand when translation is appropriate and how to structure English-rich, scaffolded instruction. This theme's strong STM alignment, combined with survey findings showing content teachers' lower confidence ($d = 0.61$) and effectiveness ($d = -0.56$) compared to ESL specialists, validates that PD gaps are both linguistically consistent and empirically consequential.

4.4. Evidence-Based Principles for Translation Use (RQ4)

To address research Question 4, we synthesized findings across three data sources: survey responses from K-12 teachers ($n = 88$), Structural Topic Modeling of open-ended responses (Q11 scaffolding strategies, Q14 translation comments), and interviews with district ESL coordinators ($n = 3$). We employed a joining display approach (Guetterman et al., 2015) to identify patterns of convergence (aligned findings), divergence (contradictions), and expansion (insights unique to one source). Table 6 presents this integration systematically.

Table 6. Joint display-integration of survey, STM, and interview findings.

Finding	Quantitative Survey ($n = 88$)	Survey STM for Open-Ended Questions (Q11/Q14)	Interview Analysis ($n = 3$)	Integration Type
Translation Prevalence	81.8% use translation; software that supports written text and/or oral (e.g., Google Translate) (89.66%), peer translation (72.2%); use of native text and bilingual dictionaries (72.72%)	Q14-T3: Bilingual app-mediated communication (30.5%); Q11-T8: Bilingual/L1 supports visuals (4.5%)	TECHNOLOGY theme (10.4%): Google Translate copy-paste (T2, 2.8%) vs. Language Line (T4, 3.5%); STRATEGIC_ESL (26.8%) includes L1 as a pedagogical tool	**Convergence**: High prevalence confirmed across all sources. **Expansion**: Interviews distinguish productive (communication tools) from counterproductive (content translation) uses

Table 6. *Cont.*

Finding	Quantitative Survey (n = 88)	Survey STM for Open-Ended Questions (Q11/Q14)	Interview Analysis (n = 3)	Integration Type
Spanish Dominance and Equity	Spanish-dominant methods; peer translation assumes shared language	Q14-T3: English Spanish bilingual focus (30.5%)	EQUITY theme (13.6%): Spanish privilege marginalizes Arabic, Vietnamese, Mandarin speakers; "15 kids Spanish, 8 are not—these 8 getting even less" English	**Convergence**: Pattern confirmed. **Expansion**: Interviews reveal inequity within the ML population, not just ML vs. non-ML.
Scaffolding Repertoire	Multiple strategies reported beyond translation	Q11: 8 scaffolding approaches—graphic organizers (19.8%), word banks (19.6%), sentence frames (19.0%), visuals (16.7%), vocabulary building (9.5%), etc.	STRATEGIC_ESL split into 3 sub-themes: native language tool (6.7%), vocabulary activities (5.9%), participation strategies (6.4%)	**Expansion**: Interviews explain pedagogical principles (scaffold vs. crutch) underlying survey-reported practices
Teacher Confidence and Effectiveness	ESL teachers reported higher confidence (M = 3.69 vs. 3.18, d = 0.61) and perceived effectiveness (M = 3.76 vs. 3.29, d = 0.57) than content teachers	Q14-T2: Content teacher reliance on tools (10.3%); Q14-T5: Newcomer support with constraints (11.5%)	PD_GAP theme (6.4%, 75% STM alignment): "Teachers say 'I need translation' when they need help modifying assignments"; MISUSE_CONTENT (11.6%): wholesale translation creates barriers	**Divergence**: Teachers perceive moderate effectiveness; leaders observe widespread misuse. Gap suggests teachers lack an evaluative framework. **Convergence**: Both sources identify the PD need
Guidelines Gap	Only 11.4% of teachers reported having district guidelines	Q14:T6: District infrastructure and parent communication concerns (18.7%)	GUIDELINES theme (21.2%, 35.8% alignment): "Teacher's first reaction is 'I'll translate everything'—that's not the expectation"; scenario-based guidance needed	**Convergence**: Gap confirmed across sources **Expansion**: Interviews detail consequences (inconsistent practice, misuse) and solutions (scenario-based policies)
Technology Double-Edged Sword	High tech use: Google Translate and other software that takes text or oral input (79 teachers, 89.66%), but confidence is only moderate (M = 3.77)	Q14:T1: Google/school resources for reading and content (9.9%); Q14-T2: Teacher reliance in content settings (10.3%)	TECHNOLOGY split: Problematic copy-paste (2.8%) vs. strategic Language Line (3.5%); "When you depend on it for whole texts, it fails you"	**Convergence**: Ubiquitous use confirmed. **Expansion**: Interviews provide an evaluative framework—appropriate for communication, inappropriate for content delivery
Scaffold vs. Crutch Distinction	Not directly measured in the survey	Implicit in Q11 scaffolding strategies; Q14-T2: Explicit crutch language—"reliant on Google Translate", "using it as a crutch", "rely on translating", "relying solely on translation"	SCAFFOLD_CRUTCH theme (10.0%, 44% alignment): Conceptual framework distinguishing temporary support from permanent dependency; "Students say 'I don't have to listen because the teacher will translate everything'"	**Convergence**: Survey teachers recognize the crutch problem ("I used to let students use translation more freely, but I noticed they were using it as a crutch, so I limit it now"), but lack a systematic framework. **Expansion**: Interviews provide an evaluative model: productive = "bridge to target language", counterproductive = dependency preventing engagement.
Misuse Patterns	Not directly measured; effectiveness ratings suggest content teachers are unaware of problems (M = 3.29, rated as "moderately effective")	Q14-T1 & T4: Teachers explicitly describe misuse—"100% reliant on Google Translate", "content teachers relying solely on Translation", "allow students to write papers in their native language. That's completely ineffective", "Google Translate. . . often have errors"	MISUSE_CONTENT theme (11.6%, 58.6% alignment): Wholesale translation, L1 literacy assumptions, broken teacher translation, double cognitive load; "Students doing double the work"	**Convergence and Expansion**: ESL teachers identify content teacher misuse patterns; leaders provide an explanatory framework for WHY these practices fail. **Divergence**: Content teachers perceive effectiveness despite ESL teachers/leaders labeling as harmful.

4.4.1. Convergent Findings

Three patterns emerged consistently across data sources. First, translation prevalence was confirmed through survey rates (81.8% use translation), STM topics capturing technology use (Q14-T3, bilingual communication 30.5%), and leaders' observation of ubiquity. Second, Spanish privilege appeared in Spanish-dominant survey methods (72.2% peer translation), STM's English–Spanish focus (Q14-T3, 30.5%), and leaders' equity concerns about multilingual marginalization. Third, the professional development gap converged across ESL-content confidence differences (M = 3.69 vs. 3.18, d = 0.61), STM topics on teacher tool reliance and help requests (Q14-T2, T9), and leaders' PD_GAP theme (6.4%, 75% STM alignment—the study's strongest convergence).

4.4.2. Divergent Findings

A critical divergence emerged between teacher perception and leader assessment. Content teachers rated translation effectiveness at M = 3.29 (moderately effective), suggesting they view translated materials as helpful. Yet ESL teachers in survey comments reported content colleagues are "100% reliant on Google Translate" and students are "Really suffering", while district leaders consistently described MISUSE_CONTENT patterns (11.6% of segments, 58.6% STM alignment). This perception gap suggests that content teachers confuse task completion with English learning—students submitted translated work, so teachers assume learning occurred.

4.4.3. Expansive Findings

Qualitative data provided two critical expansions. First, the SCAFFOLD-CRUTCH framework emerged as leaders' evaluative lens, distinguishing temporary supports from permanent dependencies. While not directly measured in surveys, Q14 comments showed ESL teacher awareness ("I noticed they were using it as a crutch, so I limit it now") and content teacher over-reliance ("100% reliant"). Leaders provided the systematic framework: productive translation builds toward English independence; counterproductive translation creates dependency.

Second, technology's dual nature was revealed: productive for communication (Language Line's 240+ languages for parent meetings, use of Google Translate for small bits of information) but counterproductive for content delivery (whole text translation using Google Translate). The survey showed high use (89.66% oral and/or text input software), but could not distinguish contexts. Leader's distinction—what you translate (direction vs. texts) and why (access vs. substitution)—provides critical guidance.

4.4.4. Seven Evidence-Based Principles

Synthesizing these patterns, we derive seven principles:

- **Principle 1: Translation as Temporary Scaffold**. Use translation for clarification and participation, then reduce reliance as proficiency develops. Avoid dependency where students disengage because they are anticipating a full translation.
- **Principle 2: Equitable Multilingual Infrastructure**. Translation support must serve all language groups, not predominantly Spanish speakers. Survey data revealed Spanish-dominant approaches (72.2% peer translation), which district leaders identified as marginalizing students speaking Arabic, Vietnamese, Mandarin, and other languages. Multi-language translation systems (e.g., Language Line's 240+ languages) ensure equitable access. When such systems are unavailable, provide non-Spanish speakers with alternative scaffolds: visual support, simplified English, or bilingual support in their languages.

- **Principle 3: Communication vs. Content Translation**. Use translation tools for family communication and directions—contexts where access matters most. Avoid wholesale translation of instructional texts and assessments, which substitutes for English engagement rather than supporting it. Content teachers rated wholesale translation as moderately effective ($M = 3.29$), yet district leaders observed it as counterproductive (MISUSE_CONTENT, 11.6% of segments); this perception gap underscores the need for clear guidance distinguishing appropriate from inappropriate contexts.
- **Principle 4: Verify L1 Literacy**. District leaders consistently noted that many MLs speak but cannot read their native languages. For these students, translated texts create cognitive burden rather than support—requiring both unfamiliar script decoding and English language development. Assess L1 literacy levels before providing translated materials; when literacy is uncertain, visual scaffolds, simplified English, and oral translation may be more effective than written L1 texts.
- **Principle 5: Modify Before Translating**. Q11 STM showed effective strategies (graphic organizers 19.8%, word banks 19.6%, sentence frames 19.0%, visuals 16.7%) precede translation needs. These modifications are particularly important for students who may speak but not read their native language, as written translation assumes L1 literacy. Visual and structural scaffolds provide support without this assumption.
- **Principle 6: Scenario-based Guidelines**. Only 11.4% reported having guidelines, despite 81.8% using translation regularly. Written policies should include concrete scenarios distinguishing appropriate use (emergency communications, family outreach, clarifying instructions) from inappropriate use (wholesale translation of assignments, substituting for English instruction, allowing L1-only assessment responses). Scenario-based guidance helps teachers apply the communication-versus-content distinction (Principle 3) in practice and addresses the perception gap where teachers view counterproductive translation as effective.
- **Principle 7: Sustained SLA Capacity Building for Content Teachers**. Content teachers need training in second language acquisition principles and modification strategies to reduce translation dependency. This need showed the study's strongest convergence (PD-GAP theme: 75% STM alignment; confidence gap: $d = 0.61$). Professional development should focus on alternative scaffolding approaches (Principle 5), the SCAFFOLD_CRUTCH distinction (Principle 1), and content-specific modification techniques. Ongoing job-embedded coaching is more effective than one-time workshops for building sustainable practices.

5. Discussion

This mixed-methods study examined translation practices among North Carolina K-12 teachers serving MLs, revealing widespread implementation (81.8% prevalence) occurring largely without institutional guidance (88.6% lack policies). The convergent parallel design enabled us to triangulate teacher-reported patterns with district leader observations, uncovering both consensus and critical tensions that inform our scaffold-versus-crutch framework.

5.1. The Scaffold-Crutch Framework: A Conceptual Contribution

The central contribution of this study is the scaffold-versus-crutch framework for evaluating translation implementation quality. Grounded in Wood et al.'s (1976) scaffolding theory and extended through Van Lier's (2004) application to language learning contexts, this framework distinguishes translation that builds towards English independence from translation that creates permanent dependency. District leaders consistently applied this evaluative lens, articulating that productive translation functions as a "bridge"

enabling meaning-making, while counterproductive translation allows students to "sit back" expecting full translation (SCAFFOLD_CRUTCH theme, 10.0% of coded segments).

This distinction aligns with Henderson's (2017) observation that effective L1 use involves strategic deployment rather than wholesale substitution for English engagement. Our framework operationalizes Henderson's insight by identifying three criteria for distinguishing scaffold from crutch: strategic application (targeted to specific learning moments), systematic fading (reduction as competence develops), and independence orientation (building self-regulated performance). When translation violates these principles, i.e., applied universally, persisting indefinitely, or bypassing English engagement entirely, it shifts from support to dependency.

The framework also resonates with Ferlazzo's (2024) question about whether translation enhances comprehension or substitutes for it and whether it encourages a growth mindset or reduces productive struggle. Our empirical findings give substance to these questions: leaders described students who disengage when they expect full translation, and teachers, in open-ended responses that acknowledged noticing students "using it as a crutch, so I limit it now." The scaffold-crutch framework provides evaluative criteria that practitioners can apply to their own contexts.

5.2. Emerging Patterns of Use

Most teachers, both ESL and content area, in our sample reported extensive use of machine-based and/or peer-translation in their classrooms with no significant differences in methods, generation approaches, or target proficiency levels. Approximately half of non-users commented on the limited effectiveness of machine-based translation. This finding challenges any simplistic narrative that positions ESL teachers as strategic users and content teachers as indiscriminate translators. The reality is more nuanced: translation has become ubiquitous across both groups, consistent with Ducar and Schocket's (2018) observation that machine-based translation represents "the key pedagogical issue of our time."

However, meaningful differences emerged in confidence and perceived effectiveness. ESL specialists reported higher confidence in translation accuracy ($M = 3.69$ vs. 3.18, $d = 0.61$) and greater perceived effectiveness ($M = 3.76$ vs. 3.29, $d = 0.56$). These medium effect sizes suggest substantive differences: ESL teachers report approximately one-half to two-thirds of a standard deviation higher in both measures. We tentatively interpret this gap through the lens of teacher preparation, though the small content teacher sample constraints definitive conclusions. ESL teachers received focused training in second language acquisition theory, proficiency progression, and principled scaffolding strategies (Janzen, 2008; Lucas & Villegas, 2013), enabling them to evaluate translation's role within a developmental framework. Content teachers, while disciplinary experts, often lack this foundation (Samson & Collins, 2012), potentially explaining both lower confidence and the perception gap.

Importantly, the small content teacher sample ($n = 22$) constrains our ability to make definitive claims about this group. The effect sizes suggest meaningful patterns warranting further investigation, but we caution against overgeneralizing from this sample to content area teachers broadly. Future research with larger, more representative samples is needed to characterize content teacher practices comprehensively.

5.3. Perception Gap

A critical finding emerged from mixed methods integration, though it should be interpreted in light of sample composition: the divergence between how teachers perceive translation effectiveness and how district leaders observe its implementation is particularly notable among the ESL-dominant sample and may differ in content-heavy contexts. Content teachers rated translation as moderately effective, suggesting they view translated

materials as helpful for student learning. Yet district leaders consistently described MISUSE_CONTENT patterns (11.6% of coded segments, 58.6% STM alignment), observing that wholesale translation creates barriers rather than removing them.

This perception gap suggests teachers may confuse task completion with English learning. As one leader explained, teachers "think, 'I did the most amazing thing. I turned the entire unit into Spanish', and then they flop it on their desk, and the student looks up like a deer in headlights." Students submitted translated work, so teachers assumed learning occurred, yet the translation may have bypassed rather than supported language development. This finding resonates with Loewen et al.'s (2020) caution that improved translation quality paradoxically increases dependency risk when teachers lack a principled implementation guide. The gap also emerged in ESL teachers' open-ended responses, where some described content colleagues as "100% reliant on Google Translate" with students "really suffering." This peer observation—ESL specialists witnessing content teacher practices—triangulates with leader concerns, and suggests the perception gap is not only a leader–teacher difference but also a variation in implementation quality that practitioners themselves recognize.

5.4. Equity in Technology-Mediated Multilingual Support

Our data indicate that current inequities in the use and availability of translation are being reified in today's classrooms. There are obvious reasons that Spanish translations would not only be better but also more easily available. Spanish is more linguistically similar to English than other languages that are not so related or share a similar alphabet. Additionally, many ESL and content area teachers have familiarity with Spanish. As a result, students who speak less commonly seen languages are at a disadvantage when it comes to translation. The translations themselves are often inaccurate or simply unavailable. When Pashto speakers came to U.S. schools, machine-based translation was almost impossible because many of the refugees did not read or write in their language, and the most common machine-based translation services did not offer text-to-voice features in Pashto.

The problem of native language literacy also impacts the Latino communities when teachers give translated texts to Spanish-speakers. Students with limited or interrupted formal education (SLIFE) do not have a high enough level of academic Spanish to adequately access the text without a more knowledgeable peer to help them scaffold the text. The translation is not enough. Moreover, a percentage of our Spanish-speakers speak Spanish as an additional language. Their native language is often an indigenous language that does not have a written form.

5.5. Professional Development and Guidance Gap

The study's strongest mixed-method convergence occurred around professional development needs. The PD_GAP theme showed 75% alignment between manual coding and STM, indicating that this need has a consistent linguistic signature across interviews. Leaders quoted teachers' explicit help requests: "Teachers say, 'I need translation,' when what they really need is help modifying what everyone else is doing." This convergence with the confidence gap ($d = 0.61$) validates that PD needs are both linguistically consistent and empirically consequential.

The finding aligns with broader research on teacher preparation gaps. Lucas and Villegas (2013) documented that few teacher education programs systematically prepare candidates to work with linguistically diverse students, and Samson and Collins (2012) found that fewer than half of teachers serving English learners have received formal training. As has been previously mentioned, there is little guidance provided that would allow both ESL and content area teachers to use translation in strategic and structured ways to help

support English language acquisition, and subsequently content acquisition. In many ways, technology advanced much, much faster than schools have been able to keep up. The implementation of translation has been piecemeal and dependent on information scattered between various groups of teachers and their communities. Very rarely has guidance been provided at a school or district level. Our data suggest that the teacher preparation gap extends specifically to translation: teachers use technology without pedagogical grounding in what distinguishes scaffold from crutch.

Leaders emphasized that building SLA capacity requires sustained support rather than one-time workshops, noting it is "far superior to talk through these issues during planning... that's where we really work through what the classroom needs." This preference aligns with research on effective professional development emphasizing sustained, contextualized support rather than decontextualized training (Darling-Hammond et al., 2017). In fact, we believe that most of the issues related to translation could be addressed with well-designed, evidence-based guidance coupled with PD to support the implementation of that guidance. This PD should be divided into several categories, including guidelines for ESL and content area teachers, guidelines for site-based administrators, and guidelines for parents who may be concerned that their children are not using an English-only model of instruction. With regard to PD for content area teachers, it should be coupled with concepts from second language acquisition to help them understand the need and use of strategic translation.

Our findings situate translation within broader post-pandemic technology integration patterns. The COVID-19 pandemic accelerated technology adoption in language education (Herrero & Spence, 2023), with translation tools becoming increasingly accessible and sophisticated. Post-pandemic, translation technology has become a permanent classroom fixture, 89.66% of translation-using teachers employ software supporting text or oral input. This ubiquity reflects Tiwari et al.'s (2021) observation that technology, once supplementary to instruction, now forms the core of delivering lessons and engaging students.

Yet, widespread adoption has outpaced the development of evidence-based implementation guidance. Only 11.4% of teachers reported having district guidelines despite 81.8% using translation regularly, a policy-practice gap that leaves implementation quality to individual teacher judgment. Leaders described having to "clamp down" on wholesale translation, noting that "a teacher's first reaction is, 'I'll just translate everything'. That's not an expectation." The implication is that technology's accessibility and its ease of use may encourage counterproductive practices when pedagogical guidance is absent.

This finding resonates with Laoha et al.'s (2025) observation that AI tools in EFL contexts require teachers to develop new competencies beyond technical proficiency. The challenge is not merely learning to use translation technology but understanding when and how to deploy it strategically. Our seven evidence-based principles represent an initial attempt to codify such guidance, synthesizing empirical patterns with expert practitioner wisdom.

5.6. Implications for Practice

Part of the strategic use of translation is helping districts know when machine-based translation is adequate and when traditional person-to-person translation is necessary. Listed below are occasions when traditional person-to-person translation (such as Language Line) is an essential component because of the high needs for real-time negotiation of meaning.

- Communication with newcomers;
- Communication with families, including instances where translation is required by law;
- Emergency situations.

We write this article as a first step in providing guidance on using translation. The seven principles that we outline are based on the data that we collected, but also reflect other relevant research in the area. They provide actionable guidance for practitioners navigating translation decisions.

1. Use translation as a temporary scaffold, reducing reliance as proficiency develops.
2. Ensure equitable access across all language groups, not just Spanish speakers.
3. Reserve translation for communication contexts (directions, family outreach) rather than wholesale content translation.
4. Verify students' L1 literacy before providing written translations.
5. Modify instruction first—graphic organizers, word banks, sentence frames, visuals—before resorting to translation.
6. Develop scenario-based district guidelines distinguishing appropriate from inappropriate use.
7. Invest in sustained, job-embedded professional development in SLA principles for content teachers.

Our principles are intended as transferable rather than generalizable. Districts can apply these principles to their specific linguistic demographics, technology infrastructure, and teacher preparation contexts. A district serving predominantly Spanish-speaking students faces different challenges than one serving highly multilingual populations; a district with robust interpreter services (e.g., Language Line) has options available to those without such resources. Our principles provide a framework for local adaptation rather than prescriptive mandates.

6. Limitations and Future Research

This study provides valuable insights into translation practices among K-12 teachers serving MLs, yet several limitations warrant consideration. First, the sample was drawn exclusively from North Carolina and relied on voluntary participation, resulting in potential self-selection bias. Teachers interested in translation or concerned about its use may have been more likely to respond. The modest sample size (n = 88) and overrepresentation of ESL specialists (75%) limit statistical power and generalizability. Content area teachers (n = 22) were substantially underrepresented relative to their prevalence in school, constraining our ability to characterize their practices comprehensively. While ESL specialists' expertise makes their perspectives particularly valuable, the sample composition limits claims about typical content teacher practices statewide or nationally.

Furthermore, recruitment through ESL-focused professional networks (university PDS system, ESL educator Facebook groups, professional association listservs) likely contributed to ESL specialist overrepresentation, potentially underrepresenting content teachers who may have different professional networks and less engagement with ML-focused communities.

Second, the study relied on self-reported data, which may be subject to social desirability bias or inaccurate recall. Teachers' perceptions of translation effectiveness and confidence in accuracy cannot be independently verified against student learning outcomes or actual classroom implementation. The divergence between content teachers' self-reported effectiveness (M = 3.29) and district leaders' concerns about misuse illustrates this limitation—teachers may not recognize counterproductive practices. Observational studies examining actual translation use in authentic classroom contexts would complement self-report data and provide more objective evidence of implementation quality.

A related limitation concerns the absence of detailed participant demographic information. To maximize response rates and maintain anonymity, our survey instrument did not collect information about teachers' educational backgrounds (degrees held, certification

pathways), professional development history related to MLs, or personal multilingual experiences. These factors could meaningfully shape translation practices and perceptions. For instance, teachers who are themselves multilingual may approach translation differently than monolingual teachers, drawing on personal experiences navigating between languages. Future research should explore how teachers' educational perception, training history, and personal language backgrounds influence their translation implementation decisions and perceived effectiveness. Similarly, the anonymous survey design precluded district-level identification, preventing analysis of how different district contexts, resources, or support structures may influence translation practices; the guidelines variable (11.4% reporting district guidelines) serves as the only proxy for institutional support differences in our data.

Third, the small number of interview participants (n = 3), while offering rich qualitative insights from experienced district leaders, limits the transferability of findings. All three leaders served in North Carolina districts and brought specific contextual perspectives that may not represent other states, rural contexts, or districts with different demographic profiles. The interview sample's purposive selection prioritized depth over breadth, appropriate for understanding expert perspectives but constraining generalizability.

Fourth, the cross-sectional design captures practices at a single time point during the post-pandemic period when technology integration was accelerating. Longitudinal research tracking how translation use evolves as teachers gain experience, as technology improves, and as students progress through grade levels would reveal whether strategic early translation supports long-term English proficiency or whether dependency patterns persist and undermine language development. The study cannot establish causal relationships between translation practices and student outcomes.

Fifth, only 10 teachers (11.4%) reported having district guidelines, precluding robust statistical comparison of practice patterns between those with and without formal guidance. This small subsample prevented us from definitively establishing whether guidelines shape practice quality, though qualitative data suggest their importance. Similarly, this study could not systematically examine how L1 literacy levels—a critical factor identified by district leaders—interact with translation effectiveness, as this information was not available for survey participants' students.

Finally, the study's focus on North Carolina limits understanding of translation use in states with different policy environments, linguistic demographics, or technology infrastructures. North Carolina's Spanish-dominant ML population may not reflect contexts serving more linguistically diverse populations. The findings may be most applicable to similar Southern states with growing ML populations and comparable technological access. Additionally, North Carolina is an English-only state with no true bilingual education. Our results may be different from a state with a stronger bilingual education component, where first language (L1) use has a history of being encouraged.

Future research should address these gaps through multiple directions. Student outcome studies linking translation approaches to English development (specifically, productive language use, academic vocabulary growth) and academic achievement are critically needed. Intervention research testing the seven evidence-based principles would establish their practical utility: Does job-embedded PD (Principle 7) shift content teachers toward strategic use? Do scenario-based guidelines (Principle 6) improve implementation quality? Multilingual equity research should investigate whether Spanish-dominant resources create disparities for Arabic, Vietnamese, and Mandarin speakers. Longitudinal studies tracking students and teachers over time could illuminate whether strategic early translation supports long-term bilingualism or whether dependency patterns undermine proficiency development. Finally, multi-state comparative research would identify how state poli-

cies and demographics shape translation practices and outcomes. This study provides foundational evidence, but the field requires systematic investigation of implementation, outcomes, and equity to guide responsible translation use in technology-enhanced multilingual classrooms.

7. Conclusions

Translation technology has become ubiquitous in multilingual K-12 classrooms without evidence-based implementation guidance. This mixed-methods study examined how teachers actually use translation, how practices differ between ESL specialists and content teachers, and how expert practitioners distinguish productive from counterproductive implementation. The scaffold-versus-crutch framework emerging from our analysis provides conceptual clarity for a practice that has been characterized by conceptual confusion, i.e., the same tool can support or undermine language development depending on implementation.

Key findings include: translation is widespread among both ESL and content teachers (81.8%), with ESL specialists showing higher confidence and perceived effectiveness; a perception gap exists between teacher beliefs about translation effectiveness and expert observations of its implementation; professional development needs showed the study's strongest convergence across data sources; and equity concerns arise from Spanish-dominant approaches that marginalize other language groups.

The seven evidence-based principles synthesized from survey patterns, STM topics, and leader expertise provide initial guidance for navigating translation decisions. As machine-based translation continues to advance in accuracy and accessibility, the question is not whether it will be used but how. Our framework and principles contribute to ensuring that translation serves its appropriate role: a temporary scaffold supporting MLs' journey toward English proficiency, not a permanent crutch preventing the very development it purports to support.

Author Contributions: Conceptualization, E.P. and T.C.B.; methodology, N.G. and E.P.; software, N.G.; validation, E.P., T.C.B. and N.G.; formal analysis, N.G.; investigation, E.P. and T.C.B.; resources, E.P. and N.G.; data curation, N.G.; writing—original draft preparation, E.P. and N.G.; writing—review and editing, E.P., N.G. and T.C.B.; visualization, N.G.; supervision, E.P.; project administration, E.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Institutional Review Board at the University of North Carolina Wilmington (protocol code Protocol #H25-0767 and date of approval 24 March 2025).

Informed Consent Statement: Informed consent was obtained from all participants involved in this study.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The study involved human participants, and the data contain information that could potentially identify participants or their affiliated school district.

Acknowledgments: The authors acknowledge the K-12 teachers and district ESL leaders who participated in this study. We also thank the school districts across North Carolina that facilitated survey distribution and leader recruitment. Their willingness to share their experiences with translation practices made this research possible.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
EFL	English as a Foreign Language
EL	English Learner
ELL	English Language Learners
ESL	English as a Second Language
FREX	Frequency and Exclusivity (term weighting in topic modeling)
IRB	Institutional Review Board
L1	First Language (Native Language)
L2	Second Language
ML	Multilingual Learner
NMT	Neural Machine Translation
PD	Professional Development
RQ	Research Questions
SLA	Second Language Acquisition
SLIFE	Students with Limited or Interrupted Formal Education
STM	Structural Topic Modeling
WIDA	World-Class Instructional Design and Assessment

Appendix A. Survey on the Use of Translation with Multilingual Learners

Thank you for considering participating in this survey on the use of translation with multilingual learners. The purpose of this survey is to examine how teachers are using translation in their classrooms as an instructional strategy. Translations have become more common as advancements in technology have made them easier to produce and more accurate.

This survey should take 5 min to complete. There are no risks associated with your participation in this study. We hope to use this data to provide guidance and information on how teachers are currently using translation in the classroom.

Your responses will be completely anonymous, and no identifying information will be collected, and your computer will not be tracked. The data will only be used for research purposes. Participation is voluntary, and you may choose to stop at any time without any consequences. By continuing with this survey, you indicate that you consent to participate.

If you have any questions, you may contact Dr. Eleni Pappamihiel at pappamihieln@uncw.edu

Your Role

Please tell us about you.

1. What is your role?
   - ○ ESL teacher
   - ○ Content area/General education teacher
2. What grade level(s) do you teach? (Select all that apply)
   - ○ Elementary (K-5)
   - ○ Middle grades (6–8)
   - ○ High school (9–12)
3. Do you use translation as an instructional scaffold in your classes?
   - ○ Yes
   - ○ No

Non-use of translation

1. You have indicated that you do not use translation in your classroom. We would appreciate it if you could tell us why. Check all that apply.
   - ○ I don't feel comfortable using a language I do not know.
   - ○ I don't feel that it is effective.
   - ○ I'm not allowed to use translation as a classroom scaffold.
   - ○ I feel that it interferes with students' ability to learn English.
   - ○ I don't feel that we should be using any language other than English.
   - ○ Other: _____

Main Questions

1. Do you allow students to use their native languages to socialize with each other in class?
   - ○ Yes
   - ○ No
   - ○ Other: _____
2. Which proficiency levels do you use translation with? Check all that apply.
   - ○ Only with newcomers.
   - ○ Mostly with newcomers but also with Level 2s.
   - ○ Anyone who feels they need it. Students choose if they want to use it.
   - ○ Anyone I feel needs it at any level. Teachers choose when translation is used or offered.
   - ○ Other: _____
3. In what ways do you use translation in your classroom? Check all that apply.
   - ○ I use an app to translate text.
   - ○ Students use translation apps as they need.
   - ○ I provide access to texts written in the students' native language.
   - ○ Students use native language dictionaries for classwork.
   - ○ Students translate for each other when they're working in groups or pairs.
   - ○ I use an app to translate spoken instructions or conversation.
   - ○ Other: _____
4. How are you doing your translation? Check all that apply.
   - ○ I speak and/or write in the language I'm translating.
   - ○ I use software or an app to translate for students.
   - ○ Students generate the translation as they need it, either in groups or on their own.
   - ○ Students use simultaneous translation devices such as earbuds.
   - ○ Other: _____
5. How confident are you in the accuracy of the translations you use?
   - ○ 1 = Not confident at all
   - ○ 2
   - ○ 3
   - ○ 4
   - ○ 5 = Extremely Confident
6. How effective are the translations in accomplishing your instructional goals?
   - ○ 1 = Not confident at all
   - ○ 2
   - ○ 3
   - ○ 4
   - ○ 5 = Extremely Confident

7. What other instructional scaffolds do you use in addition to translation?
   Enter your answer: _____
8. Does your school or district provide you with guidelines on how to effectively use translation?
   - ○ Yes
   - ○ No
9. We understand this is a very complicated issue. Is there anything else you can tell us about how you or someone you know uses translation?
   Enter your answer: _____

Appendix B. Survey on the Use of Translation with Multilingual Learners

Introduction: Thank you for agreeing to participate in this interview. We are interested in understanding how translation is being used in your district to support multilingual learners, and your perspective on effective and ineffective practices. This interview will take approximately 20–30 min. With your permission, we will audio/video-record for transcription purposes. Do you have any questions before we begin?

1. How is translation currently used for instruction by teachers in your district?
   - ○ Probe: Can you describe some specific examples you have observed?
2. What guidance does your district provide to ESL and/or general education teachers about the use of translation?
   - ○ Probe (if guidance exists): Can we obtain a copy of this guidance?
   - ○ Probe (if no guidance): Why do you think formal guidance has not been developed?
3. In what ways have you seen translation used in general education classrooms?
   - ○ Probe: What tools or methods do content teachers typically use?
4. In what ways have you seen translation used in ESL classrooms?
   - ○ Probe: How does this differ from what you observe in general education settings?
5. Do you feel that general education teachers are using translation in a way that advances English language acquisition?
   - ○ Probe: In what ways? Can you provide specific examples of effective or ineffective practices?
6. Do you feel that ESL teachers are using translation in a way that advances English language acquisition?
   - ○ Probe: In what ways? What distinguishes effective from ineffective use in observation?
7. Would you be willing to talk about this again with us after we have processed responses from across the state?

Closing: Thank you for your time and insights. Is there anything else you would like to add about translation practices in your district that we have not covered?

References

Bischof, J. M., & Airoldi, E. M. (2012). Summarizing topical content with word frequency and exclusivity. In *Proceedings of the 29th international conference on machine learning* (pp. 201–208). Omnipress.

Boneau, C. A. (1960). The effects of violations of assumptions underlying the *t*-test. *Psychological Bulletin*, *57*(1), 49–64. [CrossRef]

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. [CrossRef]

Calis, E., & Dikilitas, K. (2012). The use of translation in EFL classes as L2 learning practice. *Procedia—Social and Behavioral Science*, *46*, 5079–5084. [CrossRef]

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.

Cook, G. (2010). *Translation in language teaching: An argument for reassessment*. Oxford University Press.

Creswell, J. W., & Clark, V. L. P. (2017). *Designing and conducting mixed methods research*. Sage publications.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*(1), 7–29. [CrossRef]

Cummins, J. (1991). Interdependence of first- and second-language proficiency in bilingual children. In E. Bialystok (Ed.), *Language processing in bilingual children* (pp. 70–89). Cambridge University Press.

Custodio, B., & O'Loughlin, J. B. (2017). *Students with interrupted formal education: Bridging where they are and what they need*. Corwin.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). *Effective teacher professional development*. Learning Policy Institute. [CrossRef]

Ducar, C., & Schocket, D. H. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate. *Foreign Language Annals*, *51*(4), 779–795. [CrossRef]

Echevarria, J., Vogt, M., & Short, D. J. (2017). *Making content comprehensible for English learners: The SIOP model*. Pearson.

Ellis, R. (1985). *Understanding second language acquisition*. Oxford University Press.

Ferlazzo, L. (2024). *When is it ok to use Google Translate in the English-learner classroom?* Education Week. Available online: https://www.edweek.org/teaching-learning/opinion-the-use-or-misuse-of-google-translate-in-the-ell-classroomwhen-is-it-ok-to-use-google-translate-in-the-english-learner-classroom/2024/06 (accessed on 29 November 2025).

Garcia, O., Johnson, S. I., & Seltzer, K. (2017). *The translanguaging classroom: Leveraging student bilingualism for learning*. Caslon.

Garcia, O., & Kleyn, T. (2016). *Translanguaging with multilingual students*. Routledge.

Garcia, O., & Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. Palgrave Macmillan.

Groves, M., & Mundt, K. (2015). Friend or foe? google translate in language for academic purposes. *English for Specific Purposes*, *37*, 112–121. [CrossRef]

Guetterman, T. C., Fetters, M. D., & Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *The Annals of Family Medicine*, *13*(6), 554–561. [CrossRef] [PubMed]

Henderson, K. (2017). Teacher scaffolding of translanguaging. *Linguistics and Education*, *42*, 24–36.

Hernandez Jaramillo, A. (2019, May 13). *A historical overview of the "grammar-translation" method for teaching foreign languages. Our Language Blog*. Government of Canada. Available online: https://www.noslangues-ourlanguages.gc.ca/en/blogue-blog/methode-grammaire-traduction-grammar-translation-method-eng (accessed on 29 November 2025).

Herrero, C., & Spence, P. (2023). Introduction: Reflection on post-pandemic pedagogical trends in language education. *Modern Languages Open*, *0*(1), 31. [CrossRef]

Howard, E. R., Sugarman, J., Christian, D., Lindholm-Leary, K. J., & Rogers, D. (2007). *Guiding principles for dual language education* (2nd ed.). Center for Applied Linguistics.

Janzen, J. (2008). Teaching English language learners in the content areas. *Review of Educational Research*, *78*(4), 1010–1038. [CrossRef]

Lantolf, J. P., Kurtz, L., & Kisselev, O. (2017). Understanding the revolutionary character of L2 development in the ZPD. *Language and Sociocultural Theory*, *3*(2), 153–171. [CrossRef]

Laoha, R., Chomthong, W., & Pongpanich, W. (2025). Artificial intelligence and English as a foreign language (EFL) teachers' competencies: A systematic review. *Higher Education Studies*, *15*(3), 262–292. [CrossRef]

Loewen, S., Isbell, D. R., & Sporn, Z. (2020). The effectiveness of app-based language instruction for developing receptive linguistic knowledge and oral communicative ability. *Foreign Language Annals*, *53*(2), 209–233. [CrossRef]

Lucas, T., & Villegas, A. M. (2011). A framework for preparing linguistically responsive teachers. In T. Lucas (Ed.), *Teacher preparation for linguistically diverse classrooms: A resource for teacher educators* (pp. 55–72). Routledge.

Lucas, T., & Villegas, A. M. (2013). Preparing linguistically responsive teachers: Laying the foundation in preservice teacher education. *Theory Into Practice*, *52*(2), 98–109. [CrossRef]

Mavrou, I. (2020). Working memory, executive functions, and emotional intelligence in second language writing. *Journal of Second Language Writing*, *50*, 100758. [CrossRef]

National Council of Teachers of English. (2020). *Position paper on the role of English teachers in educating English language learners (ELLs)*. Available online: https://ncte.org/statement/teaching-english-ells/ (accessed on 29 November 2025).

North Carolina Department of Public Instruction. (2024). *English learner enrollment data*. Available online: https://www.dpi.nc.gov/districts-schools/classroom-resources/office-teaching-and-learning/standard-course-study/english-language-development/ml-identification-data (accessed on 29 November 2025).

Patterson, D., Gonzalez, J., Holzle, U., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D. R., Texier, M., & Dean, J. (2022). The carbon footprint of machine learning training will plateau, then shrink. *Computer*, *55*(7), 18–28. [CrossRef]

R Core Team. (2025). *A language and environment for statistical computing*. R Foundation for Statistical Computing. Available online: https://www.r-project.org/ (accessed on 29 November 2025).

Revelle, W. (2025). *psych: Procedures for psychological, psychometric, and personality research* (R package version 2.5.6) [Computer software]. Northwestern University. Available online: https://cran.r-project.org/package=psych (accessed on 29 November 2025).

Richards, J. C., & Rodgers, T. (2001). *Approaches and methods in language teaching* (2nd ed.). Cambridge University Press.

Roberts, M. E., Stewart, B. M., & Tingley, D. (2019). stm: An R package for structural topic models. *Journal of Statistical Software*, *91*(2), 1–40. [CrossRef]

Samson, J. F., & Collins, B. A. (2012). *Preparing all teachers to meet the needs of English language learners*. Center for American Progress. Available online: https://www.americanprogress.org/article/preparing-all-teachers-to-meet-the-needs-of-english-language-learners/ (accessed on 29 November 2025).

Savignon, S. J. (2017). Communicative competence. In J. I. Liontas (Ed.), *The TESOL encyclopedia of English language teaching* (pp. 1–7). Wiley. [CrossRef]

Snow, C., & Uccelli, P. (2009). The challenge of academic language. In D. Olson, & N. Torrance (Eds.), *The Cambridge handbook of literacy* (pp. 112–133). Cambridge University Press.

Tati, J. S., Ejus, G. A., & Askar, M. A. B. M. (2024, October 9–10). *Exploring the role of translation in enhancing ESL learning outcomes.* [Conference presentation]. 3rd International Multidisciplinary Academic Conference, Kota Kinabalu, Malaysia.

Tiwari, P., Seraphin, H., & Chowdhary, N. R. (2021). Impacts of COVID-19 on tourism education: Analysis and perspectives. *Journal of Teaching in Travel & Tourism*, *21*(4), 313–338. [CrossRef]

Torchiano, M. (2020). *effsize: Efficient effect size computation* (R package version 0.8.1) [Computer software]. Zenodo. [CrossRef]

Turovsky, B. (2016, November 15). *Found in translation: More accurate, fluent sentences in Google Translate*. Google Blog. Available online: https://blog.google/products-and-platforms/products/translate/found-translation-more-accurate-fluent-sentences-google-translate/ (accessed on 29 November 2025).

Van Lier, L. (2004). *The ecology and semiotics of language learning: A sociocultural perspective*. Springer.

Warschauer, M., & Matuchniak, T. (2010). New technology and digital worlds: Analyzing evidence of equity in access, use, and outcomes. *Review of Research in Education*, *34*(1), 179–225. [CrossRef]

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., Francois, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Muller, K., Ooms, J., Robinson, D., Siedel, D. P., Spinu, V., . . . Yutani, H. (2019). Welcome to the Tidyverse. *Journal of Open Source Software*, *4*(43), 1686. [CrossRef]

Wolf, M. K., Farnsworth, T., & Herman, J. (2008). Validity issues in assessing English language learners' language proficiency. *Educational Assessment*, *13*(2–3), 80–107. [CrossRef]

Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry*, *17*(2), 89–100. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.