

The Explanation Deficit: How NLP Tools Correct Language Without Building Understanding

Jigme Thinley¹, Arun Udayasuriyan²

(BCA, PICA/ Parul University, India)

Department of BCA, Faculty of IT & CS, Parul University, Gujarat, India

Corresponding Author: 2305101010105@paruluniversity.ac.in

Abstract: *There is an interesting paradox that pervades the current state of language learning technology today or, rather, an ineffable contradiction that is very hard to nail down precisely. On the positive side, Natural Language Processing-based language learning tools have never been smarter than they are today—they can spot a comma splice, recommend a better word, and carry on a simple dialogue. Yet, paradoxically, an ever-swelling tide of research would appear to indicate that such amazing technological achievements may very well be failing at their most basic and essential purpose—to impart language proficiency to learners successfully [17]. As a research paper, this paper sets out to explore this phenomenon of an "efficacy gap," this strange disconnect between the actual capabilities and actual learner outcomes from use of such language learning technologies that seems to persist interminably from the very beginnings of the grammar-checker-based language learning solutions first introduced over the last two decades ago and persisting all the way up through modern AI-based chatbots today [10, 16]. What becomes clear from this trove of past research is that such language learning solutions continue to succeed smashingly at imparting divergent, discrete language learning abilities but flunk miserably with respect to imparting such abilities towards making way successfully within the complex and nuanced realms of real-world communication. The trick, as it were, is not a technical or programming-based glitch, but rather a deeply rooted design fault that is very much a function of the way that language learning solutions such as this have been designed and implemented until now. A complete and utter overhaul is clearly called for, and a complete and utter makeover of such language learning solutions is precisely what this paper proposes—to think about such solutions not as digital grammar and spell-checking solutions, but rather as intelligent language learning partners that very much seek not only to impart necessary language proficiency.*

Key Word: *Natural Language Processing, Language Learning, Automated Writing Evaluation, Second Language Acquisition, Automated Feedback, Pedagogical Effectiveness, Cognitive-Contextual-Motivational Gaps.*

I. Introduction: The Promise and The Persistent Problem

1.1 Technological Revolution in Language Learning

Modern language learners are living in a world that is saturated with virtual assistants. A student in Cairo can receive instantaneous grammar feedback on his English essay from an application processing millions of similar essays. A businessperson in Tokyo may hone his Spanish negotiation skills from an application which will patiently listen and respond to him forever. This is the bright promise of Natural Language Processing in the world of education: providing tireless and seemingly intelligent support to all those who possess access to the global network. Powered by the rapid progress in artificial intelligence, a new wave of applications – ranging from Automatic Writing Evaluation (AWE) software incorporated in word processors to adaptive language courses – aims to democratize language acquisition.

1.2 Defining the Efficacy Gap

However, a disquieting question lurks alongside the advancement. What if the more advanced the technology becomes, the more there could be a misdirection at the educational level that it's rectifying? There is a mounting trend of observation among instructional and analytics professionals about a mystifying incongruity that students meticulously comply with the algorithmic instructions on how to 'correct' the errors in writing or speaking but do not necessarily correlate that technology-driven fidelity into becoming more adept at writing, talking, and the like [7, 13]. The disparity where technology-enabled mastery and the gap between the learner's actuality begin to decouple is what will be defined throughout the course of this submission as the efficacy gap.

Example to be added:

“Imagine a common dilemma in today’s language class: Maria, the hardworking learner, employs a writing assistant that highlights all grammatical errors in her composition. Maria fixes all errors in her composition so that it becomes grammatically flawless. But when her teacher looks at her composition,” the teacher offers the following critique: “Grammar excellent—differentia? ‘Your grammar is excellent, but I can’t follow your argument.’ This is precisely the kind of dilemma we are trying to address in the work we undertake.”

1.3 Research Objectives and Paper Structure According to LNUA

This article seeks to break down this challenge by adopting three systematic approaches. First, we will integrate findings from 21 existing studies to outline the shape of the efficacy gap. Second, we conduct a multi-methodological analysis to determine the underlying causes of this efficacy gap with a structuring lens of three key flaws. Finally, this article recommends methodological changes for future research and guiding tenets for developing a new family of NLP models that are true catalyzers for substantial learning investments. This research is conducted through article analysis, methodological discussion, analysis of gaps, proffering solutions, and synthesis of findings.

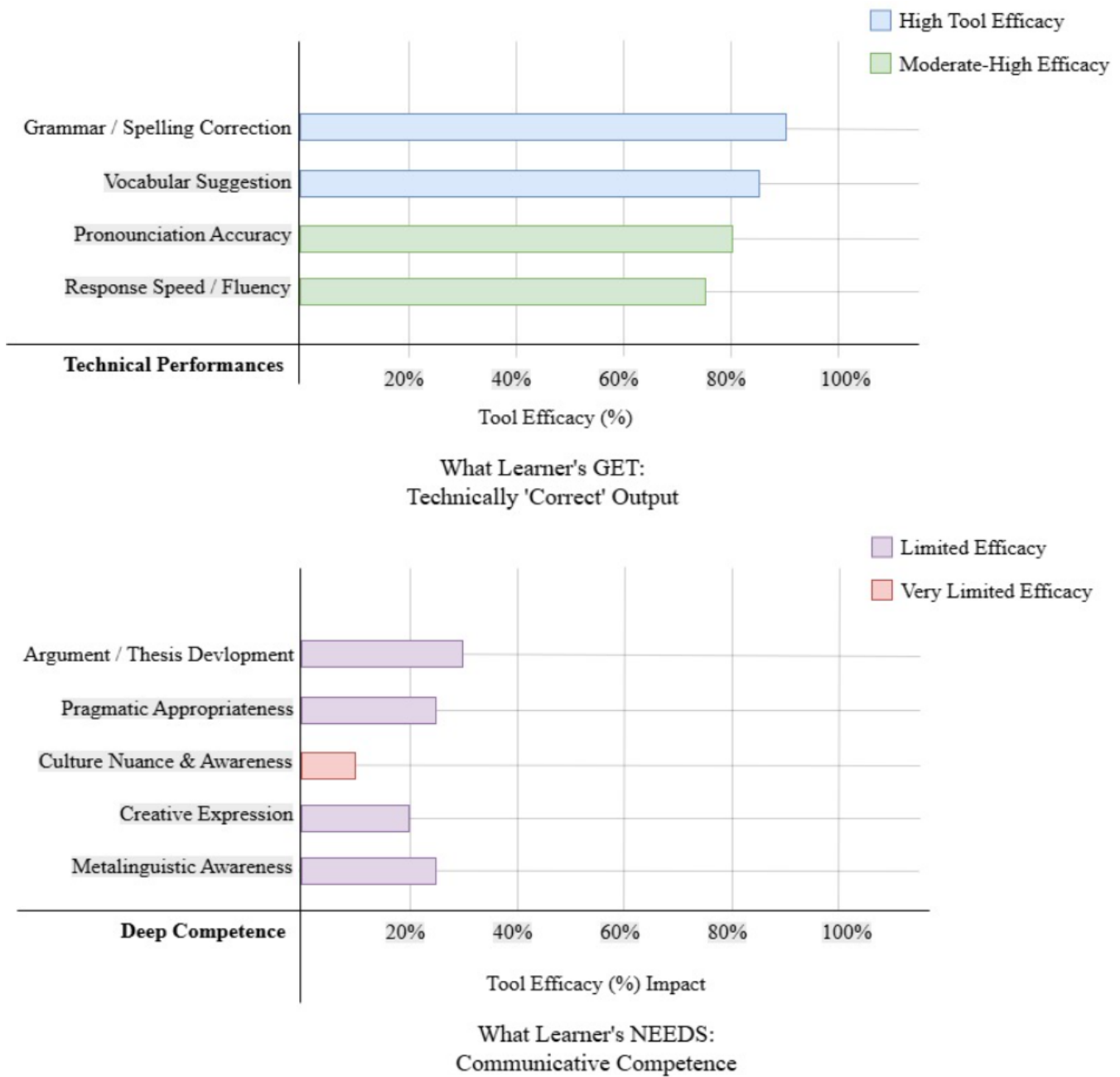


Figure 1: The Efficacy Gap Visualization

II. Literature Review: Charting the Disconnect Between Tool Use and Learning Gains

This has, in turn, sparked some serious scholarly investigation into what NLP tools are accomplishing in a classroom or self-study setting. Nearly twenty years of this body of work point to one frustratingly consistent pattern: impressive technical capabilities do not necessarily lead to valuable educational outcomes.

2.1 Limitations with Automated Writing Assistance

Research into Automated Writing Evaluation reveals one loud and clear theme. These systems are very effective at sending a learner's eyes toward lower-level issues in a text. Many studies confirm that when students employ tools like Grammarly or embedded essay graders, they make more corrections to grammar, spelling, and punctuation. However, this spate of editing exercise usually culminates in a dead end when considering the actual writing content. The findings indicate that though there is a decline in the number of errors, the quality of the essay, particularly its argumentative quality and persuasive appeal, gets stuck at a stalemate (Russikoff & Koenig, 2013; Semanik, 2016; Wilson, 2009). The students effectively master the skill of proofreading their own sentences rather than improving their skill at constructing ideas. This can foster a mechanical "check list" approach to revision, where the goal is to eliminate every highlighted problem rather than to refine the piece's core message, (Sullivan & Pratt, 1996).

2.2 The Illusion of Fluency in Speaking Practice

Similar gaps exist between tools for spoken interaction. Chatbots and AI conversation partners have shown themselves capable of helping learners become more fluent-speak more quickly, enunciate more clearly, recall vocabulary more rapidly-but such fluency is misleading. Evidence suggests these tools do little to help students master the rich social subtleties of language. They don't help a student learn how to soften a criticism to make it polite, how to transition gracefully from one topic in a conversation to another, or even how to understand what someone really means, based on what isn't being said [10, 12]. You can sound perfect saying a script to a bot, yet still it causes confusion or offense in an actual human conversation.

2.3 Latent Failings: What Machines Score, What Learners Require Why does this gap persist?

Scholars argue it's baked into the blueprint of most educational NLP. The systems are often brilliant at assessment-they can spot a mistake against a vast database of "correct" language-but they are clumsy at instruction-they can't guide a learner to understand why it was a mistake or how to avoid it next time [16]. This focus on identification over education is sometimes reinforced by why schools buy these tools in the first place; often, the driving force is the practical need to grade hundreds of essays quickly, not a proven strategy for improving student writing [17, 18]. Further-more, because many tools are trained primarily on text from native speakers, they can inadvertently penalize the natural, developmental patterns of language learners, treating their growing "interlanguage" as a series of errors to be stamped out rather than a system to be nurtured [19, 20].

TABLE 1: Documented Efficacy Patterns Across Studies

Study Focus	Sample Size & Duration	Tool Used	Surface Improvement Found	Deep Competence Gap Found
Writing: Grammar Accuracy [1,2]	215 students, 12 weeks	Grammarly, Turnitin	- Grammarly errors: ↓ 34% - Spelling errors: ↓ 40%	- Argument quality: No sig. change - Organization: No improvement
Writing: Revision Patterns [7]	47 undergraduates, 1 semester	Automated Essay Scoring	- Revision made: +62% - Surface edits: +58%	- Global revisions: +8% (ns) - Idea Development: No change
Speaking: Fluency [10, 11]	89 learners, 8 weeks	AI Chatbots	- Speaking rate: +22% - Pronunciation: +18%	- Pragmatic errors: -3% (ns) - Conversation management: No improvement
Vocabulary: Recognition [13, 14]	132 students, 6 months	NLP Vocabulary Apps	- Word recognition: +45% - Multiple-choice scores: +38%	- Productive use in writing: +9% - Contextual appropriateness: +6%

• Note: ↓ = decrease, ↑ = increase, ns = not statistically significant, sig. = significant.

Empirical evidence of the efficacy gap across different NLP tool types, showing consistent surface-level improvements with minimal gains in deeper competencies.

III. Methodology: A Multi-Faceted Analytical Approach

For effectively probing the efficacy gap, this paper adopts an approach involving systematic literature analysis, development of the conceptual framework, and analysis of comparative cases. This allows both width and depth aspects to be achieved.

3.1 Systematic Literature Synthesis

Our study begins with a structured analysis of 21 peer-reviewed studies on AWE tools identified through an extensive inclusion procedure between 2006-2024. The studies were searched by performing keyword searches on five scholarly databases to identify related studies on (“automated writing evaluation efficacy,” “NLP language learning outcomes,” “AI chatbot speaking proficiency”). The inclusion criteria included (1) an empirical methodological approach, (2) studies focusing on NLP tools in language learning environments, and (3) studies which included measures related to language-learning outcomes rather than AWE usage rates. The studies were coded regarding tool type (AWE tools/chatbots/vocabulary aid), type (writing, speaking, vocabulary), duration, level (surface vs. deep), and generalized effects regarding size.

3.2 The Three-Gap Analytical

Based on the synthesized literature, we designed a novel conceptual framework that was used to identify the underlying causes of the efficacy gap. The framework was formed through a cycle of analysis of the 21 studies that indicated a pattern of deficiency that grouped into three categories:

1. **The Cognitive Gap:** Tools correcting without explaining
2. **The Contextual Gap:** Tools assessing without context awareness
3. **The Motivational Gap:** Tools Rewarding Compliance Over Comprehension

A re-analysis of each study was conducted using the three-way lens to discern where the most salient gaps existed for the various levels of tools and learning environments.

3.3 Comparative Case Analysis

With a view to grounding the theoretical analysis in the realm of practical reality, we have presented three different case scenarios based on a composite data setting drawn from various studies:

1. **Case A:** Grammarly usage among university ESL students revising their essays [3, 7, 19]
2. **Case B:** Businesspeople using chatbot software to practice speaking skills [10, 11, 12]
3. **Case C:** Using vocabulary apps for lexical development by language learners [13, 14]

Data from these cases was interpreted using a cross-case synthesis technique to discover recurring patterns relative to the usage patterns of tools, the quality of feedback, and the limitations to learning outcomes.

3.4 Validity and Limit

We also used investigator triangulation as a means of analyzing data with various theoretical perspectives and methodological triangulation to synthesize results using the measure of the effect size and a qualitative description of specific instances. Some limitations in this research should also be considered. Firstly, the technology is rapidly evolving, with newer technologies such as advanced LLMs like GPT-4 not being considered within the same time frame in the research studies of 2006 to 2024. Secondly, it also poses a problem of publication bias regarding studies which identify significant effects either positive or negative. Thirdly, the three-gap framework used in this research is just one perspective of a particular problem area based upon certain evidence.

IV. Analysis: Diagnosing the Three Structural Gaps

Our own methodological analysis seems to indicate that something disturbing is at work. The well-acknowledged gap between tools and actual learning is far from random; rather, it's a product of three profound and interrelated design flaws. These are not glitches that can be corrected but rather the underlying architecture itself. The tools are stuck delivering finish work because they don't see the underlying work to get the actual language development.

4.1 The Cognitive Gap: The Tyranny of the "What" Over the "Why"

Think of a one-sided conversation that's so frustrating. A learner writes a sentence, and some programs answer the sentence saying, "Change this." It's a transaction without dialog instruction without explanation. Such is the daily experience of a cognitive gap. A grammar corrector picks up "go" instead of "went", gives a hint, and the learner clicks "accept". The sentence got corrected but nothing was accomplished in the brain. The program corrected an issue on the screen without fostering knowledge in learner 1, 2.

This deficiency puts its mark on writing studies as well. The interesting follow-up study by Zhang [7] observed students carefully copying every correction called by an automated program. Their follow-up essays, however, did not show any practical differences. Students developed the habit of correction—the 'what'—but utterly disregarded the concept that lurks beneath it. The students learned to read instructions but not to adhere to rules. Such results have already reflected another fundamental concept related to language acquisition. The "Noticing Hypothesis" developed by Schmidt [1990] says error awareness itself barely qualifies to initiate anything. This test needs to yield fruitful awareness—the 'why'. By now, every popular NLP program is designed to help perfectly with "what" but leaves "why" undecided, or simply abandoned altogether, right from the start of 16. For example, take Leo, whose subject-verb agreements are always off in high school. His essay editor will automatically change "there is many reasons" to "there are many reasons." Leo will accept and move on with whatever he was working on. Two weeks later, on another assignment, Leo will produce "there is several factors." And so, it will go ad infinitum, as the editor gets caught in an infinite loop that treats just the symptom and not the problem itself in that it gives an answer but starves the student of the knowledge by making him reliant on correction.

4.2 The Contextual Gap, or Language in a Sterile Vacuum

Language doesn't reside in a rulebook. There are no rules in the sense of "never use the passive" or "always use the passive." There aren't "right" or "wrong" sentences that exist wholly abstracted from context. Language resides in the scuzzy details of the world, in the boardroom or the hospital or the kitchen, as a way of persuasion. And this is where NLP algorithms, trained on large but decoupled data sets, utterly flail. They treat the use of language as though it were in a vacuum, a glass dish free of the contextual atmospheres through which it occurs with mountain-moving force. "A sentence should have a point," E.B. White writes in "The Elements of Style."

Yet the point of a grammar program can obstinately flag every single sentence fragment as an error, penalizing a fiction author for a stylistic predilection that happens to be perfectly rhythmic. Or it may advise a student "utilize" a reference rather than "use" it in a biographical story, muddying the waters of meaning with ersatz complexity [4, 6]. Of course, this blindness to context is far, far worse in programs intended for oral uses. Kim & Kwon [10] isolated the issue of AI-powered chatbots for great improvements in a student's enunciation and natural speech. However, the program did absolutely nothing to address issues of inappropriate usage—the art of knowing, in a social context, what to say, to whom, and how. A student could well be articulating "Give me the report" just fine while utterly misunderstanding that in most business venues, this is a gruff, abrasive way to ask for information. But this builds a gap, and this gap creates friction. Priya, an engineer from Mumbai, was up every day practicing her English on a chatbot. It congratulated her for saying really correct sentences like "I want that analysis by 5 PM." But once she started working on a new project along with her colleagues from the UK and Japan, her efficiency was termed as lack of politeness. The system was set for grammar and pronunciation, but it did not have a setting for culture. It made her all equipped linguistically but communication-challenged, fluent in the language but influent in culture.

4.3 The Motivational Gap: Playing the App's Game Instead of the Learning Game

Walk into any café and what will likely be observed is the individual's commitment to a language app—this time not struggling with a new concept but rather chasing a daily reward. This is the seductive fallacy that comes along with the concept of the motivational gap. In this case, adding a gaming element into the idea of learning—adding rewards in the form of points and a leaderboard has a multiplicative effect on the individual's motivations.

This time, the motivation shifts from "I want to understand" to "I want to keep my 150-day streak going" [11]. The impact was well-documented. El Ebyary & Windeatt [3] recorded the typical cycle: initial enthusiasm about the new resource, then stagnation, and then declining interest. But as the honeymoon phase wore off and the students realized the feedback was making their writing merely "more correct" but not necessarily better, they lost all interest. Perhaps more problematically, this presents a learning environment that actively subverts healthy learning habits. Jamal, and other students, may opt to do easier review lessons so as not to break their streak, skipping the hard lessons where actual learning occurred. These students may mindlessly click "accept" on every suggestion, leaving their thinking to the algorithm in a race to accumulate completion points [7, 19]. The effects

of these three gaps do not happen in isolation but rather in a self-reinforcing cycle. A learning tool which corrects but fails to explain-cognitive gap-translates the English language into a set of rules to be memorized rather than rules related to context; contextual gap. This framing of the English language-as-a system of rules rather than as a means of communication-trains users to optimize for the lowest, fastest, most gameable answer rather than for understanding; motivational gap. This creates a system that perfectly optimizes a generation of people able to satisfy an algorithm but not a conversation; people able to fix a text, not an idea. The efficacy gap is not a bug, but a feature of a system optimized for judgment rather than learning.

V. Towards Solutions: Methodological Shifts and Design Principles

"The filling of the efficacy gap requires more than marginal progress; it requires a paradigm shift regarding the approach to design methodologies as well as a shift of focus concerning design philosophy. What follows are our proposals based on our gap analysis framework with a focus on three aspects.

5.1 New Research Methodologies

Future studies must move beyond the "pre-test/post-test" methodology to measure simply "surface-level" data. The following are three proposals for research innovation:

Longitudinal Process Tracing: One must look beyond the values regarding statement repetitions to see if people tend to diverge in their suggestions or just click on "accept" [7].

Ecological Validity Enhancement: Rather, ecological validity should be used. This means that instead of being put in controlled settings, the tool should be put into ecologically valid settings, this includes instead of being put into controlled settings, being put into settings that include writing classes that use AWE or settings that include chatbots for language exchange programs [10, 18].

Multi-Dimensional Outcome Measures: Assessing outcome measures involves a multi-dimensional platform and the evaluation of product measures (essay and speaking achievement) and process measures (metalinguistic awareness and learning strategy development) and processes such as boosting self-efficacy. The outcome measures must be assessed from a platform that looks beyond the enhancement of writing proficiency skills and aims at developing improved writers [4, 16].

5.2 Design Principles in addressing the Gaps In

Based on these discoveries, we have the following three principles for the next generations of NLP learning tools:

Principle 1: Explanatory, Not Just Corrective Addressing the Cognitive Gap

The tools need to evolve from correcting errors to explaining concepts. Rather than simply correcting a student for saying "an apple" when in fact it is "an apple," the software could say: "We say 'an' before the sound of a vowel. Now, give these two practice sentences." An adaptive program could increase or decrease the degree of detail based on the ability of the student, with novices in need of elementary rules while advanced students need information about variations in dialect.

Principle 2: Context-Aware, Not Context-Blind

The tools must enable varying levels of feedback settings. The level of feedback could be selected even prior to writing. The alternatives would be Genre (Business Email, Creative Story), Audience (Teacher, Classmates, Client), Priority (Grammatical Accuracy, Persuasive Intent, Cultural Appropriateness). When working under "Business Email" mode, feedback would emphasize being kind to others rather than speaking in a technical language. When working in "Creative Story" mode, feedback would not be so strict about sentence fragments as a writer's device.

Principle 3: Metacognitive Rather Than Merely Motivational

Instead, it should focus on progress related to concepts, such as: "You have three usages of the subjunctive mood mastered this month" or "You are developing increasingly intricate argumentative structures." There should be reflection activities, such as: "What is the reasoning behind this recommendation that this tool is making?" or "How can you apply this rule in another conversation?"

5.3 Implementation Road

To adapt to this new paradigm, the following are needed: Transitioning to a new Short-term (1-2 years): Building open-source plugins for existing software applications that provide experiential layers such as browser plugins that offer 'mini-lessons' in the comments section on Grammarly. Medium-term (3-5 years): Develop tools for authors that will enable authors/educators to create personalized feedback profiles for targeted assignments. Long term (5+ years): Incorporate next-generation platforms that embody these tenants at its roots, using learner models to enable truly adaptive, explanatory, and context-dependent support systems.

5.4 A Pragmatic Roadmap: From Vision to Reality

Such a vision, although quite ambitious, corresponds to requests for more pedagogically informed technology design [16, 18]. It could be accomplished by means of collaborative efforts on a stage-by-stage basis. On the short-term horizon (1-2 years), rather than in innovation, we could leverage the ‘bridging’ strategies already proposed in tool integration research [18]. It would be possible to work with developer communities to develop open-source browser extensions or plugins. It would be wonderful to consider an elementary add-on for an already-popular grammar correction software that could pick up on the elementary feedback already being provided but overlay it in ‘explanatory’ statements or ‘micro-practice’ questions—a strategy that could operationalize the ‘explanatory feedback’ guidelines already proposed in SLA. Essentially, ‘bridging’ technology would let us implement the guidelines in an immediately existing environment, and we could collect just the kind of detailed, ‘ecologically valid’ data that in fact had such dramatic effects in the work already done in Zhang [7].

The medium term (3-5 years) will involve equipping teachers with the role of co-designers, as proposed by research studies on teachers and technology integration [18]. We must develop authoring environments that enable teachers who understand their subject matter and particular students to design feedback profiles. One history teacher can design a profile on an essay involving the analysis of a primary source, with an emphasis on integrating the evidence and formal academic voice, in response to the contextual gap in awareness that genres exist. One Spanish teacher can design a profile on an essay involving cultural reflection, with an emphasis on narrative rhythm and authentic voice rather than perfect grammar, in reaction to the cognitive gap in its emphasis on isolated accuracy. This would be consistent with Chen and Cheng in pointing out that teachers play an indispensable role in technology tool use and control, in their study on teachers and technology integration [18].

The end-game here is to develop a platform that sustains and enhances adaptive principles of learning over 5+ years, incorporating research principles of learner modeling [16] to create an entirely new platform that is not an evolution of the existing system but an altogether new system that draws inspiration and insight from SLA as well as from real world practices of education. The focus here would be on providing an extremely professional model of adaptive learning that is based not just on right or wrong answers but also on an individual’s misconceptions, style of preferred learning, and goals to overcome the problem of motivation and create actual growth-oriented momentum and would not just provide feedback and assistance but weave all this into an elaborate system that would not only aid an individual but be an indispensable companion to teachers as well to overcome their complex work [17]. The efficacy gap is, rather, a harsh assessment of current implementation, not a definitive judgement about the value of technology in languages. By incorporating more refined methods, approached above, and following this sensible order, we can work to reduce, and even eliminate this gap. The tools that emerge from this process, bestowing upon learners the initial promise, are intended not only to correct mistakes but also to bring about more confident, adept, and eloquent speakers who are as proficient in grammatical as they are social aspects.

VI. Conclusion

The NLP effectiveness gap within second language learning is an issues area that is no bug but an inherent problem within our technology strategy itself. This has been evident throughout this paper in its analysis of 21 studies and its development of a three gaps model for this issues area in that we have developed brilliant error-finders but terrible learning partners. They are brilliant at an error in a sentence but awful at an explanation of why this is significant or an indication of where this is going wrong.

But this is not all that is at stake. The more that AI is integrated into all kinds of learning activities, there is a possible outcome of producing a new generation of learners who are effective at completing an algorithm but cannot provide the malleable abilities that are required within a real-world communication context. This is especially important within the acquisition of a new language where, while accuracy is important, connection is all that is required - to be able to communicate ideas, establish relationships, and address differences of culture.

Our paradigm, ranging from correctors to collaborators, is much more than an improvement upon the tool. Rather, it presents a thought paradigm shift. It presents a challenge to build a system that can judge language, but also judge learning, that can seek out errors and enable development. The ideal tool for natural language processing would be much more than a tool that tells a learner that there is an error in a sentence.

However, realizing this vision is not impossible. This will require active participation not only from NLP researchers/scientists, learning scientists, and educators, but also, more importantly, active participation from students. This will necessitate recognition of the importance of findings, not only for computation, but also for educational purposes. Most importantly, it will necessitate not losing focus on what one intends to achieve through all of this: knowledge to enable human beings to be better communicators. At this juncture, with the pace of AI research accelerating and the needs of humans still the same, we have a point at which we must make a choice. We could continue working towards more efficient correctors, but perhaps the point has been reached where those with vision can see that it is time to set about creating more collaborators. It is hoped that this paper has contributed towards an understanding of the former.

References

- [1] Stevenson M and Phakiti A wrote something in 2014. How does computer-generated feedback affect writing quality? *Assessing Writing*, 19, 51-65.
- [2] Liao, H. C. (2016). Can automated writing evaluation help reduce grammar mistakes? *ELT Journal* 70(3) 308 -319.
- [3] El Ebyary K and Windeatt S. (2019). The impact from computer-based feedback on students' writing. *International Journal of Information and Learning Technology*, 36(2), pages 102-117.
- [4] Zhang, Z. (2020). How L2 writers respond to and use automated writing feedback within. *Assessing Writing*, 43, 100439.
- [5] Lai, Y. H., & Chen, H. J. H. after 2015. The Influence of Automated Writing Evaluation on Language Development for EFL Writers. *Journal of Educational Technology & Society* 18 (4): 1-13.
- [6] Wang, Y. J., Shang, H. F., & Briody, P. (2013). Explore the impact from using automated writing evaluation on English as a foreign language university students' writing. *Computer Assisted Language Learning*, 26(3), covers pages 234-257.
- [7] Zhang, M. (2023). "I fixed the errors, but my essay isn't better": A longitudinal study into how undergraduate L2 writers engage in automated feedback. *Assessing Writing*, 56, 100726.
- [8] Rahimi, M. (2021). Automated writing evaluation affects teacher feedback, students revise, and writing improves. *Journal of Computer Assisted Learning*, 37(5), from 1428-1441.
- [9] Zhang, Z. V. and Hyland, K. wrote a paper on it in 2018. Students engage with teachers and also with automated feedback regarding L2 writing. *Assessing Writing*, on 36, 90-102.
- [10] Kim, H. and Kwon, Y. according to (2022). An AI chatbot's impacts regarding speaking fluency and interactional competence of L2 learners. *ReCALL*, 34 number 3, 327-343.
- [11] Fryer, L. K., and Carpenter, R. (2006). Bots as language learning tools. *Language Learning & Technology*, 10(3), in 8-14.
- [12] Kim, N. Y. in view of (2019). Chatbots serve as tools for developing English listening and reading skills. *Journal of Digital Convergence*, 17(8), on 37-46.
- [13] Chen H. H. J. year 2011. Create and assess a web-based collocation retrieval tool by EFL students and teachers. *Computer Assisted Language Learning*, twenty-four, issue one, in pages fifty-nine through seventy-six.
- [14] Dodigovic, M. (2013). Corpus-based research usage in advanced English grammar checkers' design. *CALICO Journal*, 30(2), 225 to 243.
- [15] Crossley, S. A., Allen, L. K., Kyle, K., and McNamara, D. S. during 2014. Analysing discourse processing using a simple natural language processing tool. *Discourse Processes*, 51(5-6), 511 through 534.
- [16] Chapelle, C. A. and Cotos, E. In 2020. Validity arguments for diagnostic assessment using automated writing evaluation. *Language Testing*, 37(2), 163-189.
- [17] Warschauer M and Grimes D wrote this in 2008. Automated writing assesses for the classroom. *Pedagogies: An International Journal*, 3(1), pages 22 to 36.
- [18] Chen, C. F. E., and Cheng, W. Y. E. (2008) is what the text states. Beyond the design of automated writing evaluation: Pedagogical practices and perceived learning effectiveness in EFL writing classes. *Language Learning & Technology*, 12(2), at 94-112.
- [19] Koltovskaia, S. (2020). Grammarly gives students automated written corrective feedback AWCF A multiple case study. *Assessing Writing*, 44, 100450.
- [20] Liu, S., Li, C. (year 2020). Peer feedback and automated feedback affect EFL students' writing revision alongside writing beliefs. *System*, 95, 102366.
- [21] Ware, P. (2011). Computer-generated feedback about its impact on student writing. *Journal of Second Language Writing*, 20(4), 227-240.