

Future of Vocabulary Learning

Anuj Menta

Nivedita Ganesh

Rishabh Yadav

Contents

Broad Mission

1. What is Vocabulary Learning?
2. Why English?
3. Language Learner Segmentation

Pedagogical Survey

4. Teachers
5. Academicians
6. Best Way to Learn a Language
7. Practical Learning Process

Interviews

8. Industry leaders
9. Students
10. Academic literature

Problem Definition

11. Domain
12. Specific sub-problem

Market Analysis

13. Context
14. Total Addressable Market
15. Serviceable Addressable Market

How Can We Improve?

16. Personalized Learning
17. Multimodal Learning
18. Conversational Learning
19. Leveraging Native Language

Pioneering a New Approach

Broad Mission

English is a behemoth, the likes of which the world has never before seen. With astonishing influence worldwide, it has become aspirational: a **golden ticket** to the global vistas of commerce, business, and education; a discordant divider twixt the haves and have-nots.

“Vocabulary enables us to interpret and to express. If you have a limited vocabulary, you will also have limited vision and a limited future.”

We hope to discover and propagate the most effective way to expand one's arsenal of vocabulary, and strengthen their access and ability to deploy it.

1. WHAT IS VOCABULARY LEARNING?

Vocabulary learning has traditionally been understood as learning a word and its meaning. We want to redefine it to mean:

*word + meaning + ability to use it appropriately
in speaking and writing*

Students find vocabulary learning hard because it is difficult to learn unfamiliar words and their meanings in bulk. Traditional vocabulary learning techniques do not provide sufficient information on how to use the words in correct contexts, morphology and syntax.

2. WHY ENGLISH?

We see the future of vocabulary learning not just as a superficial task, but an opportunity to allow individuals to expand their thinking and learn how to **communicate powerfully, effectively, naturally and correctly.**

With English as the lingua franca of the world, being fluent is not just an added comfort but a stark differentiating factor in many arenas. One in four people can speak or understand English, and India alone has 125 million fluent English speakers (12% of the population).

Fig. 1 is from a study in India which depicts a 76% increase in average hourly wages for men and 82% for women, if they were able to speak English fluently.

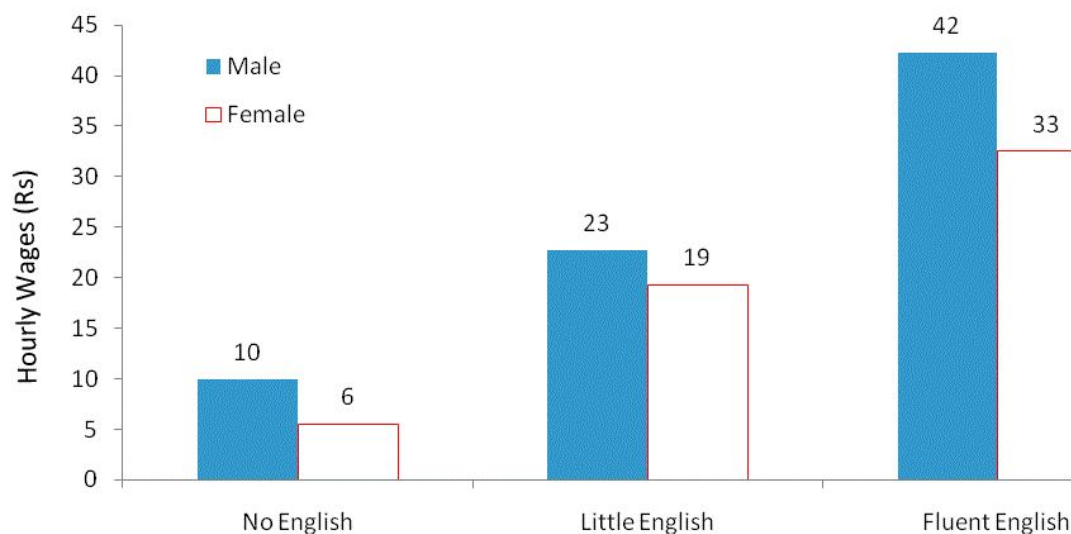


Fig. 1 Unadjusted mean hourly wages by level of English - Forbes, 2010

This could be attributed to a variety of factors; the fact that in India there is a positive correlation between higher education and fluency in English, or that fluency in English is concentrated around urban areas which typically have greater costs of living and therefore higher average wages. However, no one can deny that there is a **glass ceiling** in the form of a **language barrier**. It has become essential to be able to communicate efficiently in English to be able to rise through the ranks.

Faster career growth also occurs if one is confident and able to achieve both **clarity of thought and speech**. Without fluency in the chosen medium of communication, one is limited by their arsenal of vocabulary to a lower socio-economic status.

3. LANGUAGE LEARNER SEGMENTATION

We divided our market of language learners quite naturally based on skill; basic, intermediate and advanced. Below we explore what the biggest problems are for each of these segments:

Basic

Main use case: COMMUNICATION

- Spelling and confusion between words
- Pronunciation and enunciation

Intermediate

Main use case: EDUCATION

- Inability to retain complicated words and their meanings
- Difficulty using correct syntax and grammar

Advanced

Main use case: BUSINESS

- Ineffective usage of vocabulary at disposal
- Expression limited by depth of vocabulary

To address these problems, academicians recommend different tactics for each segment. Basic learners can best cement their spelling and articulation with traditional methods such as **practice and repetition**. This is what is commonly reflected in vocabulary learning resources today.

For intermediate learners, methods such as flash cards and fill in the blanks are being currently employed. However, it is recommended that both for them and advanced learners, the optimal way to understand correct grammar and syntax and employ vocabulary effectively is to immerse them in organic content using multiple mediums i.e. allow their **brain to absorb natural links between words**. Whilst flashcards have evolved from their

original state, to include memory retention techniques such as spaced repetition, it still provides a very shallow understanding of the word.

To ensure it is at the level of the student (intermediate or advanced), one only has to choose the most difficult resources which they can understand and work forwards from there. If the percentage of unknown words used is too high, it can be overwhelming to understand the meaning and context.

Pedagogical Survey

We explored the current state of pedagogy by consulting teachers, academicians, and academic literature, to arrive at the optimal approach to language learning.

4. TEACHERS

We spoke to teachers: across language learner segments, across geography, from schools to British Council to private tutoring, and collated the key points that resonated.

- The number one thing that had an overarching theme was that students must be kept relaxed, engaged, and motivated. To do this, one can use relatable and interactive content. Using literary prompts that **stimulate the senses** also helps to tap into creativity by making them more aware of their body and how it interacts with its surroundings.
- In general, **storytelling** is an effective way to seize a student's attention. Contemporary techniques such as text stories or comic strips have had traction due to the brief nature of their form. If one is able to tap into humour and current trends, there is sufficient academic literature to support that it is a powerful way of **learning memorably**.
- It is important to take note that teachers have difficulty tackling the **"fear of the blank page"**. To reduce this, the only way is to get students to start writing. Even if the content or delivery is bad, it is important to begin somewhere and to at least give oneself a chance at progressing

towards quality. It is also the best way of testing spelling and deeper understanding of syntax.

- Offering **bite-sized information** can also help to make sure that students are not overwhelmed. With the advancement of personalized learning, it is becoming easier to meet the student where they are and support them to achieve their personal goals.

5. ACADEMICIANS

After consulting language and pedagogy experts in academia, we were able to gain insights on the science behind education technology and how to apply it effectively.

- It is necessary to achieve the correct balance between the **receptive** (reading and listening) and **expressive** (speaking and writing). Teachers in a classroom setting usually achieve this by using the **scaffolding technique**: first demonstrate or model the idea, then leave the student to work on problems and support them as required. Reflecting this process online can in fact be even more effective as you can tailor the experience for each student so as to strike the perfect balance.
- To customise the learning experience optimally, use a two pronged approach to asking questions. First ask the question and receive their answer, then ask the student to **rate how confident** they feel about their answer. This not only serves to make the student more self aware of how judiciously they are assessing the correctness of their answers, but also as a metric for the personalized learning algorithm to **adapt** and modify the learning approach optimally.
- **Game based learning** can be more productive. Effective learning games¹ usually include: conflict, strategy, chance, aesthetics, theme, story, and rewards. If carefully designed, it can in fact enhance the learning process as it encourages **higher order critical thinking** and can utilise a competitive edge to arrest the attention of the student as well as provide external incentives.
- A method to emulate the **community** present in a classroom that is often lost online, would be to create a modified social network

¹ <https://elearningindustry.com/5-game-elements-create-effective-learning-games>

mechanism. This not only creates a forum in which students can interact and study with each other, but also to set collective goals and be held **accountable** for them by providing internal incentives.

6. ACADEMIC LITERATURE

After speaking to multiple teachers, academicians, and industry experts, we were instructed that understanding academic literature is crucial to the problem at hand. Numerous conferences around the globe witness a diverse range of talks and papers on understanding the pedagogy behind different forms of learning.

We handpicked popular papers which we could most relate to the problem, for which the findings are as follows:

Acquired vs. Learned

‘Comparing and Contrasting First and Second Language Acquisition’:

Implications for language teachers which talked in detail about the most intricate differences between first and second language. It is quite amusing how the paper refers to the first language as ‘acquiring’, since as a child, one doesn’t have any means of communication or expression and learns their first language in natural bits and pieces. Infants generally learn their first language between the ages of 0 to 6 years and most of the ‘acquiring’ phase happens in that particular age group. Learning a second language usually requires particular motivation whether it be travel, certification, recreational or emigration. The paper highlighted that emulating the natural first language learning process could be the crux of improving the efficiency of the current second language learning process.

Word to World Mapping

Adults differ from children in terms of both cognitive capacity and world knowledge. Correlated cues make otherwise-difficult categories and dependencies easier to learn. Some of the best case studies done in the past understand the behavior about nuances from the difference between the first and second language and understanding the contrast between different interactive methods of learning.

The importance of word to world pairing was emphasised. A paper on how difficult language learning is talked about a case study in which students were shown an actual duck when the word duck was mentioned and the students related naturally to the meaning of the word. Experiments on similar lines talked about teaching the meaning of 'see/look' to someone who's blind.

Nuances of Learning Across all Languages

Noam Chomsky proposed a linguistic theory that argues that the ability to learn language is innate, distinctly human and distinct from all other aspects of human cognition. Universal grammar (UG) is the genetic component of the language faculty. The basic axiom of UG is that a certain set of structural rules are innate to humans, independent of sensory experience i.e. whilst languages might have different vocabularies and grammars, **humans are born with an understanding of the laws of grammar** and have a natural predilection to learn and use them.

If we understand and formalise this structure, learning new languages and translating between them would be considerably simpler. UG theory is not fully developed, however, he has found support crossing between the older languages from around the world.

Chomsky's hypothesis was in opposition to the beliefs held by behavioural psychologists previously whom assumed that language and grammar were learned solely by listening to it being spoken. Since the common belief has now transitioned to be aligned with Chomsky, our approach to language learning must also adapt. In particular, armed with Chomsky's "The Minimalist Program", literature at the forefront of generative English grammar, we can incorporate his findings on the rules of grammar to develop new and innovative ways to transition between languages.

7. BEST WAY TO LEARN A LANGUAGE

The brain is an extraordinary machine which subconsciously does a lot more than you are aware of. When reading or listening, you are doing more than just comprehending the words at face value. In fact, you are constantly **collecting statistics** about which words appear next to each other and how syntactical structures are formed. This means your brain can anticipate quicker what is coming next, and therefore process speech more expeditiously.

Neuroscientists² have been able to use electroencephalography to study how your brain processes language, and have noticed a particular signal (the N400) change depending on whether words make sense in context or not. It is relatively small for habitual word groupings such as "milk and biscuits" but increases for unexpected combinations such as "milk and tyre". Therefore, we can measure how people respond to language, and analyse their expectations of **how words occur together**.

In language learners, the N400 changes based on proficiency. It is now understood that the better a person is at a language, the closer their N400 behavior is to a native's. This illustrated the importance of **flooding one's brain with natural content**, to assimilate collocations and other information. Experts recommend using a variety of mediums such as visual (reading),

² <https://www.fluentin3months.com/best-way-to-learn-a-language/>

auditory (listening), and audio-visual (watching). Doing this from a variety of sources and genres will provide a holistic understanding.

8. PRACTICAL LEARNING PROCESS

Here we outline what academics consider an optimal method to entrench vocabulary, its meaning, and its correct usage, whilst flooding the user's brain with organic English content.

The first step would be to decide which type of standardised English to adopt. Then, ascertain which words are being targeted and tag them all with the following information (can be found in either Google's corpora or the Advanced Learner's Dictionary of English) to create a **word profile**:

- Definition(s)
- Etymology
- Synonyms
- Genre
- Register (formality)
- Core or restricted (if role changes upon context and placement)
- Different use cases in multiple mediums (audio, visual, audio-visual)
- Contexts/ collocations

To make the word profile appeal to all kinds of learners, content should be scraped from a wide variety of high interest and **good quality natural context sources** from a myriad of domains.

The second step should be to group words and **create stories** or scenarios with them. Allow the user to interact with it by reading, reviewing, and marking difficult words which hyperlink to the word profiles. Then take the difficult words out and look at each of their meanings in context, teach the **grammatical frame** it fits into and allows the user to **observe nuances**. To do this, we can ask the user to retell the story, either by starting with nothing or by using the cloze procedure³. This will help identify **subtle mistakes** such as "succeeded to earn" (incorrect) vs "succeeded in earning" (correct).

³ <https://www.chino.k12.ca.us/cms/lib8/CA01902308/Centricity/Domain/2243/Cloze%20Procedure.pdf>

Contrary to popular conditioning, we were strictly advised to **keep flashcards for last**. The brain does not work well without context, and this should only be reserved for either last-minute learning, testing, or revising difficult word profiles.

Interviews

9. INDUSTRY LEADERS

Below are observations from experts in the industry, including successful language learning companies, NYC **Department of Education** and popular EdTech companies from India:

- **Native content** plays an important role in the learning process since it makes the content more relatable to the user.
- **Multimodal learning** plays a crucial role in improving user engagement and word retention, as students also confirm.
- Phonetic exposure to unfamiliar and foreign sounds is challenging
- Hands-free learning is becoming increasingly popular with millennials since they have a lot of time to kill during commute or during daily chores which they usually spend listening to music or podcasts.
- It is very important to understand how **academic literature** talks about learning a second language or vocabulary since they might help us understand the pedagogy behind the same.

10. STUDENTS

Student interviews were a very important part of our observations since they would be our major focus group and we needed to understand the nature of their problems. We interviewed a diverse set of students who have encountered a **significant amount of vocabulary learning**. A few overarching problems include the difficulty of natural usage of a word and unavailability of multiple examples or occurrences of a word.

Other observations include:

- Need for **community** whilst studying for added motivation.
- Personalization is not just specific to tailoring the content as the user keeps learning, but it is also trying to determine where the user starts learning. An example would be trying to **differentiate** the learning processes of someone who has been working for over 10 years against a senior year undergraduate student.
- Learn the **application** of new vocabulary rather than begin with the meanings. Notably, most other products have not addressed it.
- Students use a **combination of methods** to strengthen their learning process and it differs from person to person. This illustrates that there is no single mode of learning which is the most effective or preferred.

Problem Definition

11. DOMAIN

Users of the language learning market derive motivation from a variety of sources. For **recreational, touristic and aspirational** users a big problem is that the motivation is often short-lived. A study showed us that traveling to a new country is one of the most popular reasons for learning a new language followed by improving career prospects. A subset of this market is the **global English learning market** which has seen some innovative startups build smart tech for the market. Despite this, language learning products express difficulty in gaining return users, limited mostly by the lack of sustained motivation.

As outlined in our introduction, there are a multitude of reasons people flock to learn English - including but not limited to necessity. Case studies have been conducted to understand why and how English is the key to the door of unlimited opportunities. Pearson⁴ points out how **different countries** are embracing this practice to increase **employment** and **trade** opportunities to further grow their GDP. However, out of over **1.5 billion** people who speak English (which accounts to about 20% of the world's population), only **375 million** speak English as their **first language** giving rise to certain disparities which can be observed in the workspace and society in general. The English learning market alone stands at a whopping \$3.2Bn growing at a very fast rate.

⁴ <https://www.pearson.com/english/about/gse/case-studies.html>

Hundreds of thousands of people are hungry to **master English**, whether it be due to travel, emigration, certification or recreation.

12. SPECIFIC SUBPROBLEM

With many companies building products in the English learning space, we decided to focus our analysis on the **vocabulary learning** market to understand how people approach it for different purposes, ranging from standardized testing to learning to equip effective communication.

As outlined in the market segmentation, we can broadly classify a general English speaker into basic, intermediate and advanced.

After speaking to industry experts, academicians and students, we decided to focus on the **intermediate educational test-taking market** for the following reasons:

- Test takers are **goal-oriented** which removes the problem of motivation
- The educational test taking market is **well-defined and organized**
- The **scope for improvement** is greatest for the intermediate segment as the test taking market majorly comprises of those who are accustomed to the language but need to learn and use new vocabulary appropriately.

Market Analysis

13. CURRENT PRODUCTS

We analyzed successful companies in the educational test-taking market and compared and contrasted them based on the following parameters:

1. Usage of flashcards as a part of the learning process
2. Mobile application to provide easy on-the-go access
3. Customized assessment practice that adapts
4. Usage of fun and relatable multimodal content
5. Personalized learning tailors the approach to provide for each student.

Features	Magoosh	Princeton Review	Kaplan	Manhattan Prep	Economist Prep	Edwin
Flashcards	✓	✗	✗	✓	✓	✗
Mobile App	✓	✗	✓	✓	✓	✓
Customised Practice	✓	✓	✓	✗	✗	✗
Multimodal Learning	✗	✗	✗	✓	✓	✓
Personalised Learning	✗	✗	✗	✗	✗	✓

Fig. 2 Comparison of features offered by different companies

Magoosh

Magoosh is an online test preparation platform that helps students to prepare for examinations like GRE, GMAT, TOEFL, SAT, ACT, Praxis, LSAT, MCAT, IELTS. They make use of audio-visual content and also have subject

experts to help students with specific doubts. As a part of their test preparation, they have separate products for vocabulary learning.

Magoosh offers a free vocabulary learning mobile application which helps the users improve their vocabulary for their respective test. The application divides the words which are pertinent to the test according to its difficulty. It provides flashcards for the respective words which include the meaning along with a practical example. It employs a spaced repetition learning algorithm to ensure the word gets embedded in the mind of the user.

Princeton Review

The Princeton Review is a college admission services company offering test preparation services, tutoring and admissions resources, online courses, and books. The company has more than 4,000 teachers and tutors in the United States and Canada, and international franchises in 14 other countries. It offers test preparation services for GRE, SAT, LSAT, TOEFL, etc.

Notably, Princeton Review, unlike its fellow competitors, does not offer a mobile application for accessible vocabulary learning.

Kaplan

Kaplan educational offers services to colleges, universities, corporations, and businesses, including higher education programs, professional training and certifications, test preparation, and student support services. Kaplan offers intensive English learning and preparation programs in Europe and in the Asia/Pacific region and also provides products for standardized test preparation like GRE, GMAT, LSAT, etc.

Manhattan Prep

Manhattan Prep is an American test preparation company. It focuses on preparation for the GMAT, GRE, and LSAT. In December 2009, Manhattan Prep was acquired by Kaplan. Manhattan Prep offers both on-site and virtual preparatory programs for students worldwide. Its curriculum is designed around on a "content-based" approach that teaches the material covered on the exam rather than solely focusing on test-beating strategies.

Economist Prep

The Economist Prep provides an online interactive test preparation platform for standardized tests only for GRE and GMAT.

Edwin

Edwin is an AI-powered service for learning foreign languages. It combines the latest technology, professional English tutors, and effective educational content to better prepare for the exam. Also, its AI builds an optimal course for every student. They can practice at any time, anywhere. Moreover, it makes sure students master challenging parts and skip right over easy ones, so they prepare for the exam as quickly as possible.

The global online learning market was valued at \$190 billion in 2018 and it is expected to cross \$300 billion mark by 2026⁵. Moreover, the global language learning was valued at \$9.16 billion and it is expected to be around \$20 billion by 2026⁶.

The key factors that are favoring the market growth are:

1. Flexibility in learning,
2. Low cost,
3. Easy accessibility,
4. Increased effectiveness by animated learning,
5. The escalation in the number of internet users and growing access to broadband pooled with mobile phone and online capabilities.

Development of education technology, and the growing absorption of it into academic institutes and society is core to the burgeoning eLearning market.

14. MARKET DEFINITION

The digital language learning market for English which was valued at \$3.24 billion in 2018 is estimated to reach \$11.4 billion by 2027⁷. As of today, 1.5 billion

⁵<https://www.globenewswire.com/news-release/2019/03/06/1748554/0/en/E-Learning-Market-to-surpass-300bn-by-2025-Global-Market-Insights-Inc.html>

⁶<https://www.verifiedmarketresearch.com/product/global-online-language-learning-market-size-and-forecast-to-2025/>

⁷<https://www.prnewswire.com/news-releases/global-digital-english-language-learning-market-report-2019-market-was-valued-at-us-3-24-bn-in-2018-and-is-anticipated-to-account-for-us-11-44-bn-by-2027--300830765.html>

people speak English and it is expected to increase to 2 billion by 2020⁸. The following factors led to the growth:

- Increasing adoption of the digital education system
- Implementation of the advanced-level English curriculum in schools
- An increasing number of immigrants for higher education are the key factors driving the global digital English language learning market.

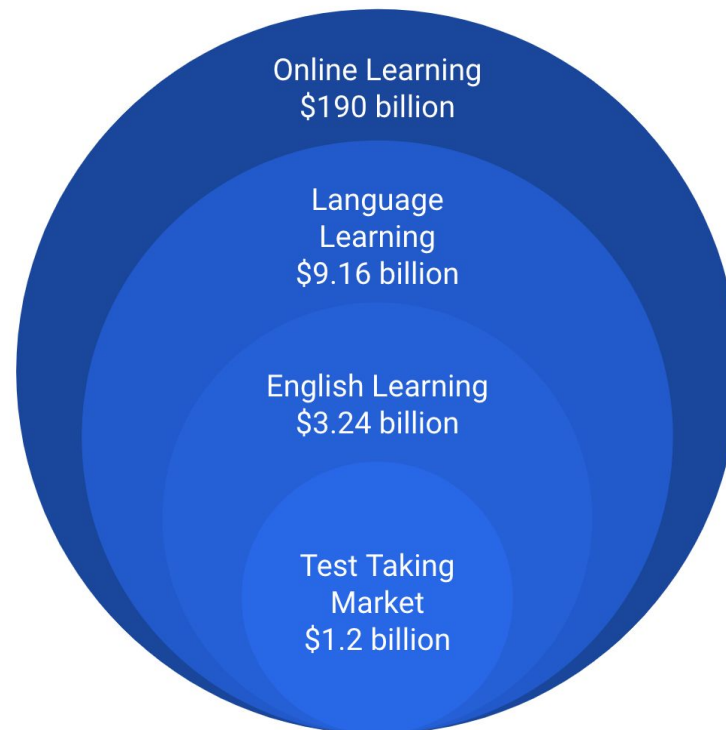


Fig. 3 Overview of the market

According to a report by IBISWorld the education test preparation industry in the US is valued at \$1.2 billion⁹ however, this is expected to increase to \$8 billion by 2023¹⁰. Concerns about the quality of public education, which have historically driven demand for tutoring as a supplement to primary education, is expected to increase. Since educational testing is the gateway to higher education, parents and students are more willing to invest time and money.

⁸ <https://www.thoughtco.com/how-many-people-learn-english-globally-1210367>

⁹ <https://www.ibisworld.com/industry-trends/specialized-market-research-reports/business-franchises/personal-service/tutoring-test-preparation-franchises.html>

¹⁰ <https://www.technavio.com/report/test-preparation-market-in-the-us-industry-analysis>

15. TOTAL ADDRESSABLE MARKET

In 2019 the test preparation industry in the US was valued at \$1.2 billion. However, it is expected to increase in subsequent years. Every year in the US 1.9m people appear for the ACT, 1.7m for the SAT, 416,000 for the GRE, and 84,600 for the GMAT¹¹. More than **15 million** people take educational tests every year.

16. SERVICEABLE ADDRESSABLE MARKET

Approximately 84,600 people appear for GMAT in the US every year and the market is estimated to be around \$8.3 million. Similarly every year 325000 people in the US appear for GRE and subsequently, the market size for GRE is estimated to be around \$60 million.

How Can We Improve?

We analyzed the feedback of more than 25 aspirants of various standardized tests regarding how the process of vocabulary learning can be improved and tried to couple it with insights provided by education technology industry experts and leading academicians in the domain of language learning.

We arrived at the following value propositions:

17. PERSONALIZED LEARNING

¹¹ <https://wsvincent.com/the-evolving-online-test-prep-market/>

Personalized learning capitalizes on students' almost instinctual ability to use technology, but it is so much more than technology and algorithms. It is the purposeful design of blended instruction to combine face-to-face teaching, technology-assisted instruction and student-to-student collaboration to leverage each student's interests for deeper learning. To aid the students it is necessary to add relevant content and examples that help them to better learn the process which is missing from all the current platforms which are available in the market as of now.

18. MULTIMODAL LEARNING

The more you immerse yourself in the language you want to learn, the more effective the learning process is. Associating a concept or a word to a situation improves the learning rate by leaps and bounds. We need to provide students with a platform that associates words with a particular instance in a TV show/ movie/ talk show/ podcast/ editorial etc. We can enhance this technique by utilizing relatable content that the user is already likely to be exposed to.

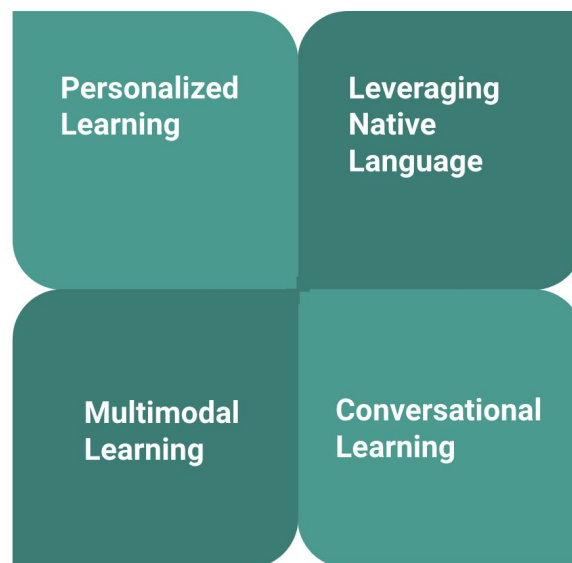


Fig. 4 Value Propositions

19. CONVERSATIONAL LEARNING

The majority of students mentioned that they had difficulty in remembering the meanings of large corpora of words. To increase retention of meanings, it is advised to frequently employ the word in various contexts. Daily usage of vocabulary that needs to be learned, in both spoken and written mediums, will ensure they are embedded in the user's lexicon.

20. LEVERAGING NATIVE LANGUAGE

As of now, no product leverages the user's native language. It would be interesting to explore how personalized learning can be augmented with the understanding of the languages within the user's scope to provide content that can act as a catalyst for the learning process.

Our vision is to create a product so effective that it transcends the user's preparation for their test, and they revert to the holistic experience in their journey to expand their lexicon and improve their hold on potent communication.

Pioneering a New Approach

We want to aid English language learners to absorb and retain large corpora of new and difficult vocabulary and help them employ it appropriately in speech and writing. Starting with formal test takers, we hope to immerse them in organic content to inculcate a native's comfort with the language whilst making the learning process entertaining and effective.