Self-Audit Your Skills Like a Ninja: An Approach to Stimulate Critical Use of Digital Tools

Ellen HEAD | Associate Professor, Miyazaki International University

FUTUREd, Volume 1, Issue 1 (2026)

Pages: 103 - 111

ISSN: 2760-8263 (print)
ISSN: 2760-8271 (Online)

Keywords:

critical thinking, spoken fluency, AI coach, self-evaluation, goal-setting

Corresponding author(s):

Ellen Head
ehead@miu.ac.jp

CC BY

Abstract: Student motivation often plateaus at upper intermediate level. Digital tools which are available easily online, allow students to get feedback on various quantitative aspects of their performance and set achievable, quantifiable goals. This article describes how a class of upper intermediate and advanced ESL students exploited various apps or sites to measure aspects of their performance such as reading level, reading speed, vocabulary size, spoken fluency and pronunciation. They were required to keep handwritten logs of what they discovered about their performance and submit handwritten reflections on the activity. It is suggested that this activity may help to raise students' awareness, of both their own level, and how various language-learning sites and apps are designed. This can be termed a "ninja approach" to online learning, in which the learner decides what they want to find out, selects the app and grabs the output with a screenshot, as far as possible avoiding tracking or logging in. The sites range from the newest AI-assisted speech coach to simple data-base-driven vocabulary size measures. In the context of increasing AI-mediated control over various aspects of language learning, this combined analogue/digital, student-centered approach seeks to hand agency back to students.

1. INTRODUCTION

Students often become demotivated because they cannot see whether they are improving (Richards, 2008). A sense of mastery is necessary for motivation (Williams and Burden, 1997). External tests do not always help because in the short term, the test algorithm relies on inferencing which may not capture how each individual is progressing. Therefore, asking students to log quantifiable aspects of their performance can be interesting and motivating for them. Tasks which harness this include timed writing, timed reading and vocabulary size tests. Starting to notice quantifiable aspects of their performance gives some control back to students, and is often novel and fun. How long can they speak for without a break? How many words can they say in one minute? And so on. Recently the power of technology to measure and track all aspects of language competence has increased greatly, meaning that students can access a range of tools for gathering quantitative data about their own output with accuracy and ease. In this article I would like to explore a "ninja approach" to the use of these powerful tools. The characteristics of this approach are that the learner sets their own goal, finds out what they want to know with minimal surrendering of their data such as email and credit card number, and keeps their own audit trail

through analogue (pencil and paper) notes and screenshots. It is possible that an app might help us more if we decide to make an account. However, my experience of teaching with apps for vocabulary, extensive reading or grammar practice, reveals that the apps have a certain addictive quality, which I think may lead to cognitive offloading (Gerlich, 2025). Students keep their heads down and focus on the app for long periods of time. There is plenty of research to show that apps work well for language learning, but very little that focuses on the possible negative impact if decisions about what to learn next become restricted to the choices in the app.

If we want to encourage critical thinking, it is important to ask "What did this app do for you?" and "How much were you able to get without logging in?" "Is this app the best one for this purpose?" and even "Were you able to break it?" instead of fostering compliance with an app. In this way we can promote learner agency and those moments of "friction" which promote learning (Hawkinson, 2025, p. 183). Hawkinson opines that "agency requires intentional friction" and the interpolation of decision-making points. (ibid, p.79). Hawkinson writes that human agency is preserved by a certain amount of necessary friction. "The friction must be productive, challenging students in ways that develop capabilities rather than merely creating obstacles." (Hawkinson, p. 183).

By designing a task to be completed "ninja style", we are getting students accustomed to some useful practices such as evaluating various tools, training them to navigate fluently between analogue and digital modes, and raising awareness that digital tools should be at their service rather than the other way round.

2. THEORETICAL BACKGROUND

In spite of the widespread excitement around advances in AI in relation to language education, some ELT professionals see a significant risk to critical thinking, autonomous learning and academic literacy (Baldrich, Perez-Garcia, Santamarina-Sanco, 2025). Olsson (2026) goes further in stating that a certain amount of slow and painstaking work or "drudgery" is actually a necessary part of developing academic literacy. Thus, there is a well-developed rationale for creating tasks which encourage students to keep offline notes and to use digital tools selectively for particular purposes, with critical awareness.

Hawkinson (2025) describes several approaches to the issue of AI use, from "passive consumption" in which the users surrender their thinking capabilities to AI, to "human-in-the-loop" in which humans maintain a higher degree of control over when and how AI is used. Here is Hawkinson's definition of human-in-the-loop:

> Human in the loop maintains high human agency while AI capabilities remain relatively basic. Students and teachers actively direct AI tools that respond to prompts but do not plan independently. This scenario requires conscious institutional effort to preserve agency even as it forgoes some benefits of more advanced systems. (Hawkinson, p.100)

In the "human-in-the-loop" scenario, inefficiency is seen as necessary part of the learning process.

AI tools do tend to recommend themselves. In their current iterations AI will rarely say "The best thing you can do is go off line now." This appears to undercut student self-efficacy and autonomy. As studies by Zimmerman and others show, self-efficacy requires students to be able to estimate their own capability and to think of their own next actions in their learning sequence. (Zimmerman, 1995, p.204).

3. THE SELF-AUDIT EXPERIMENT

3.1. Participants in this study

The participants in my advanced communicative English class in Spring 2025 had CEFR levels ranging from B1 to C1. Over half came from what Kachru (1992) would term "second circle" countries where they had done a significant amount of their secondary school education in English. They were in the third semester at a small university in Japan, studying for a BA in liberal arts in English. The communicative English class used the textbook World Link 4. The textbook was

somewhat easy for them. The course requirements included the completion of some assignments on the proprietary learning management system, which students were able to ace easily and, in some cases, were a little bored with. Those from English speaking countries in particular were complacent about their level because it was much higher than that of the Japanese students with whom they interact during EMI classes.

In connection with the approach advocated here, it is relevant to note that digital tracking had been a topic of discussion in the class already, thanks to a unit in the textbook, World Link 4 (Hughes, 2021). Data privacy is discussed in Unit 4 "The Second Self", which includes questions about social media image and a reading about "How to disappear." Students were very interested in the issue.

3.2. The Flow of the Class

The module on "skills self-audit" was integrated over four weeks of the class. Table 1 below gives an overview of the components of the activity.

Table 1 - Classroom Activities Week-by-week

Week	Activity
1	Teacher introduces the self-audit activity.
2	Self-audit of spoken fluency (Yoodli)
3	Follow up to self-audit of spoken fluency: analogue record-keeping
4	Self-audit of reading speed (Ace reader, Spreeder) Vocabulary size tests (Lex tutor, Vocabulary Size test)
5	Follow up to self-audit; reflection, discussion, goal setting. Upload work to moodle.
6	Teacher feedback on self-audit task.

Near the start of their third semester, students were asked to watch a TED talk by computer programmer Varun Puri, designer of an AI speaking coach app called Yoodli. Puri made the app to help himself because he was not a confident speaker. After having students watch the video at home, I showed students my own result from speaking Yoodli. After having a good laugh at the fact that Yoodli analytics under "conciseness" noted that I spoke 41% more than necessary, I showed students how to find and try an evaluation for free under the presentation tab. (There is also an interactive sales interview with an AI interlocutor, which some students tried.)

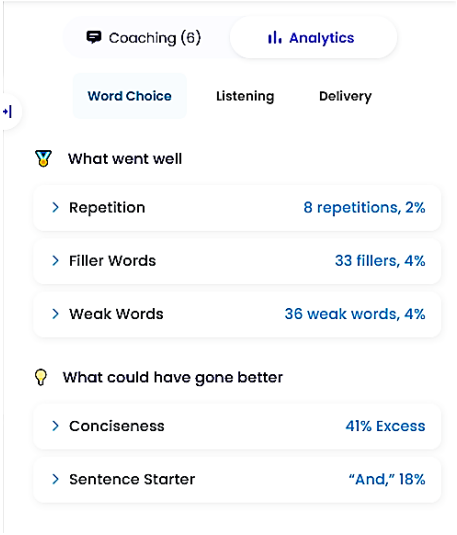

Figure 1 - AI Speaking Coach Feedback (Ellen's speech)

The feedback is extraordinarily detailed, relevant and human-like. I asked students to try out the app and to save screenshots and fill in some notes of what they discovered, on a paper worksheet which I handed out. The format of the worksheet was very open-ended rather than following the categories in Yoodli, because I wanted students to choose for themselves what they found noteworthy from the Yoodli feedback. Yoodli requires a log-in but it is possible to have five sessions for free. It can offer feedback on the basis of only 30 seconds of speaking. I encouraged them to play with it and see what they could learn, saving the output as a screenshot and pasting it into MS word which could be uploaded to the class moodle to share at least a portion of the output with me. At the end of the class, I gave out a paper pro-forma, telling them to keep notes about what they did and what they thought of it. I explained that this assignment had several parts and would be graded not on the level they attained but on completing the assignment and keeping a record carefully.

In the next class I showed students the online reading site Spreeder, which allows you to paste some reading material into a text window and then deliver it in small chunks at a particular speed. Usually if you start at about 60 words per minute, students can read it but you can ask them what speed they want, and increase until they cannot read aloud any more but can still follow cognitively. There are many things they can learn from this: most people do not realize that in L1 they read faster than they speak, and in the case of these students many read faster than they thought. They also gained awareness of the trade-off between speed and depth of understanding. However spreeder is less satisfying to use alone because the user has to input text to put into the window, meaning that it is not completely unseen text. So we looked at another reading test site, Ace reader, which allows the user to find a speed independently, selecting from 13 levels and also selecting the theme and choice of stories. As before, students had some time to play with the site in class and were asked to follow up at home. This time many students tried to do the test during class and scored rather low due to the distraction of the others around them. A few people tried again at home and improved their speed. The take-away was that their reading speed would vary depending on the level of interest they are feeling, the difficulty of the text among other factors.

The last quantifiable measure we investigated was vocabulary size. The idea of vocabulary size was familiar to them from selecting graded readers on X Reading. I gave a mini lecture about how the computer makes inferences, based on the words that you know, to estimate the words you know which are in the same frequency range. Students may not realize this kind of mechanism underlies much objective testing of language skills. In this class, it seemed that they really liked learning new words. I asked them to choose from two vocabulary size tests, or do both of them if they wanted. Many chose lex tutor. The interface is a classic computer style from 20 years ago. In spite of the old-fashioned interface, the algorithms underlying the test are very sophisticated and are updated regularly. In the initial versions, databases were used to determine word frequency and infer the total number of known words (Cook, 2004).

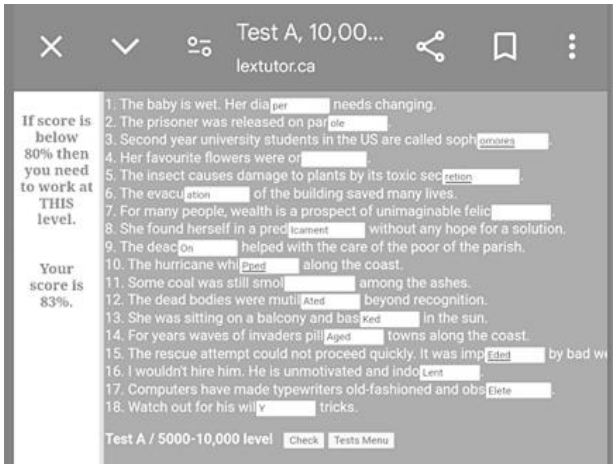

Figure 2 - Lex Tutor Vocabulary Size Test

Going through the process of selecting a test from a particular frequency level, should help students to understand how such tests work, at least I thought so. However when students selected the level of lex tutor, as before, some chose the 2000 word level in spite of my warning that that they must know 6,000-14,000 words just to be able to take the class. Others were highly motivated and tried both lex tutor at the 10,000 word level and the vocabulary size test primarily aimed at native speakers, which is based on recognition and adaptive technology. At larger vocabulary size levels, the accuracy is debatable but given that some students had finished their secondary education in English it was not impossible that the app diagnosed two of them in the expert native speaker range (30,000 words).

3.3. How did students use the self-audit?

In this section we will look at how several students used the self-audit activity. The students were among the more confident members of the class and readily agreed to share their data for the purpose of this publication. Figure 1 shows the proforma which students were given on which to report on their results. It has spaces to record the website used, results, and their own reflection.

They were allowed to upload as many files to the moodle tab as they wanted, and just select the highlights to write about on a paper reporting form as seen in figure 3.

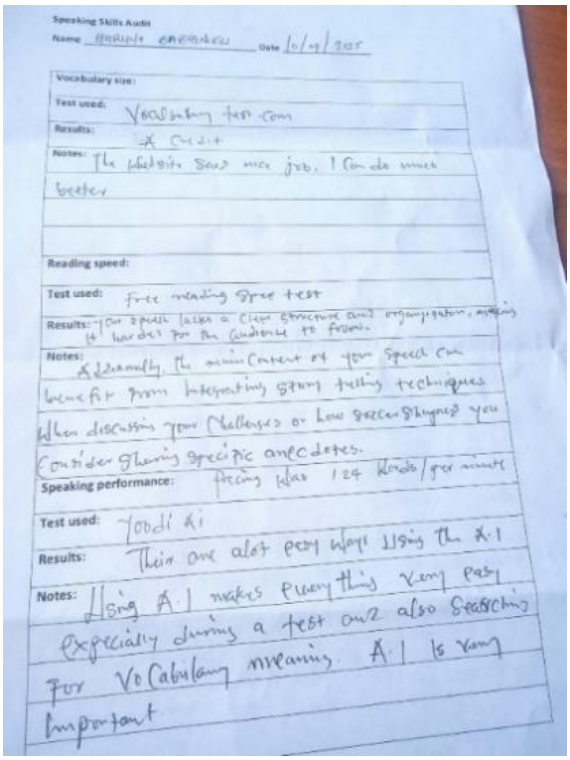

Figure 3 - Worksheet for analogue record-keeping (Student C)

Student A used analogue record keeping only. It was interesting that he tried to use advanced vocabulary in his record. He also seems to have done the reading speed test several times so as to get a higher score. Here are his records.

Skills Audit

Name: Student A
Date: April 10, 2025

Vocabulary size

Test used: Vocabulary Size.com

Results: My vocabulary size is 10,000.

Notes: Activity included plethora of complicated words that I memorized constantly when I faced it.

Reading speed:

Test used: Ace reader

Results: 320 words per minute

Notes:

I increased my speed from 249 to 320 and kept my accuracy. I think it is very beneficial exercise that helps us develop our skills.

Speaking Performance

Test used: Yoodli: 113 words per minute

Notes: I had few words that I repeated several times and started with such as so, like, umm. However, it was beneficial for me by showing not only weak points but also strong for example AI recognized my speech, as well as said my "sentences were well constructed".

Student B lavished great care on the self-audit. He provided three screenshots from Yoodli, one from the vocabulary test and two audio files of his own spoken reflection on the use of the apps. On the worksheet he copied salient parts of the feedback from the app. In his written and spoken reflection he provided evaluative comments about the apps. which I quote in full. The written comment was not extensive but his spoken reflections were thought-provoking. This learner was very experienced as an AI user having worked as a developer before coming to university. His reflection raises some relevant problems with using AI as a speech coach.

> And also, for me, communication is more about how we talk with each other in every day. So, from the eye sight or the body language, that will affect a lot about the daily conversation. So, I still prefer my human partners for my speaking section more than an AI.

His comment on the vocabulary size test also highlighted his feeling that he had not done himself justice. This student was from an EFL background and therefore scored highly on vocabulary tests which were related to the coursebook as opposed to the broader based vocabulary size test. He found it a bit irritating.

> And I found that a lot of words that I know about it, it means when I can, when I saw that word, I know it's that word, I know the meaning, but when we do that kind of test, I just don't figure it out what you're meaning in time. But after I can recognise a lot of words, but yeah, it's just not fit my

> style. That's all.

He included a somewhat defensive comment on the AI's evaluation of his vocabulary:

> Second file
>
> because that's just a quick talk, so I cannot express a lot of my vocabulary. And also, personally, when I do the practise with my friends, they also said my vocabulary skill is quite impressive.

This part seems to flip back to talking about the vocabulary size test

> But I'm not usually use that [kind of word] in a daily life communication. So let's go to the another part is the reading speed and presentation pacing. So in that day, I think I speak kinda calmly.
>
> Okay, so as you can see the result in the picture, they said I should go for faster. But when I talk like in something I haven't prepared, it's quite hard to speak at a fast speed because my brain is not progress the story quick enough. Okay, that is my problem. So yeah, maybe write something down [before speaking?] and no, that is not. Sorry, that's not cool, I guess. (Comment from the audio file of spoken reflection by student B.)

Student C wrote across the worksheet ignoring the spaces for the reading and vocabulary size and focusing on the feedback from Yoodli. He copied out several parts of the salient feedback from Yoodli and uploaded 8 screenshots to show the feedback from Yoodli and other aps. He did not analyze his feedback in detail but writing down the salient point allowed us to communicate about it effectively when chatting later.

Skills Audit

Student C

Speaking Performance

Test used: Yoodli:

Notes: (Copied from Yoodli screen) The main content of your speech can benefit from integrating story-telling techniques. When discussing your challenges or how soccer shaped you, consider telling specific anecdotes.

Reflection: there are a lot of easy ways using the AI.

Student D uploaded a single word document with 10 screen shots pasted into landscape format and a photo of the handwritten reflection. He used more of a note-taking style. By including a screenshot of his speech contents, he communicated his objections to using AI for speaking coaching:

> If you ask me, I'd say no, it's not a good solution for participants practicing speeches. As you can see, I'm very terrible at this. Many people will try to use this but it's always good to have human feedback than AI feedback and that is all for now.

Skills Audit

Name: Student D
Date: April 10, 2025

Vocabulary size

Test used: Vocabulary Tester.com

Results: My vocabulary size is 31,661

Notes: According to the test it is C2.

Reading speed:

Test used: Free Reading Speed Test Ace Reader

Results: 162 words per minute

Notes: The website says I can do better but I don't know about that.

Reflection

The website was easy to use and the options were not too shabby or limited.

Speaking Performance

Test used: Yoodli:

Results: 149 words per minute

Notes: There were a lot of criteria for this test/audit. AI assistant shows that there is much work left to be done.

Student E uploaded 8 screenshots saved into a pdf along with his handwritten comment uploaded as a photo. He showed his agency clearly by connecting the self-audit with his own language learning goals and priorities.

Skills Audit

Student D

Speaking Performance

Test Used: Yoodli

Word choice – not bad

Listening – good

I got some pronunciation corrections from the site and a few tips on how to get better and follow up questions I should expect.

Reading speed 168 words

Free reading test. Com.

Comprehension

75%

I need to increase my speed 33% more. The site has great stories and similar to x reading.

Vocabulary

I got C2 The site is quite easy to use but it does not provide corrections. C2 is a very good level so I am satisfied.

My goal henceforth would be to get 270 to 350 recommended reading speed. Although I want to achieve that by practising on my own pace as I have more important private goals. Reading speed might be ok but understanding comprehension is what helps in passing exams. From April to May I would like to improve in speaking/pronouncing, since the yoodli app made a lot of pronunciation corrections.

Vocabulary is not my priority but I will add a few to the cart.

4. DISCUSSION

Students used the self-audit in their own way and focused on their own goals. Allowing them to choose their own parameters for the levels in the reading test and lex tutor made the assessment less accurate but it gave them more control. The measures were quite different from what I would have expected given my own estimation of the students' performance. However, if we consider this activity as a fact-finding mission in which students see what they can get from various tools, then what is interesting is their level of critical engagement with the various tools. Comments such as "I still prefer my human partners" and "vocabulary is not my priority but I will add a few to cart" showed that students were doing the activity with a high level of awareness of their own needs and goals. In some cases, students copied the feedback from the app, and it was hard to tell whether they had thought much about it. One person even miscopied the speaking feedback into the space intended for reading and vocabulary, and left those blank. This shows the importance of following-up with a teacher or learning advisor in order to get the maximum benefit from a ninja approach. This kind of data seemed to be too personal for sharing through peer coaching, even though students were happy to talk about whether the apps worked or not and help each other in the beginning stages.

Students were surveyed eight months after the course to see what they remembered about the activity. Interestingly, the vocabulary size test and the reading speed test were remembered as well as the more sophisticated app and seemed to be equally popular. One person liked the feeling that they were learning "crazy" advanced words and had taken both vocabulary size tests. Having a feeling of being able to check their level was said to be helpful. Several people said that improving their vocabulary would help them to improve their speaking.

Students were impressed by the speed and human-like nature of the feedback from Yoodli. Most of them were already using Siri or other voice-driven apps but had never experienced evaluation from an app. Yoodli provided a reality check for the most

confident members of the class and also gave positive feedback to those who were less confident. Half mentioned getting feedback on their speaking as useful because it helped to reduce their nerves or helped them to know their level.

Asked "which tools would you use again", they were equally divided between the vocab size, reading and speaking apps. Five out of eight said they used other apps regularly for their independent study. Quizlet, the vocabulary size test and Duolinguo were mentioned and it seemed one person had used Yoodli again too. One mentioned using other apps such as Duolinguo to learn Chinese.

As mentioned above, the students had discussed data tracking in the context of online privacy, in an earlier module. In the follow-up survey I asked students if they were concerned about tracking by apps. On the contrary, one person felt tracking would lead to personalized learning and higher efficiency. In response to a question about tracking within the apps they use for class such as X Reading, two people said they find it useful to check their reading speed and one to check whether they are spending too much time online.

5. CONCLUSION

The self-audit experiment was intended to motivate students by raising their awareness of their strengths and weaknesses and assist them in refocusing their goals. To some extent it seemed to fulfill that purpose. In this activity, students are asked to use various different tools rather on a single integrated one, and to use them with as minimal tracking as possible. This activity highlighted old fashioned, off-line note-taking as a means of keeping our own awareness and control of our progress. It was suggested that students could be encouraged to take an opportunistic attitude to the use of digital tools in order to model criticality rather than compliance. This was a helpful way to build students' awareness of educational measurements which they could do themselves. Language learning systems for online practice need to serve students' goals rather than becoming digital drudgery in which they have to hit some targets set by a teacher. It was encouraging to see that students had their own goals which they could monitor themselves and evaluate independently of exams. It is hoped that the approach showcased here will be helpful to others who are interested in supporting students with goal setting and staying motivated, and with critical thinking in relation to AI use.

ACKNOWLEDGEMENTS

The contribution of the students who have allowed their data and reflection to be shared in this article is gratefully acknowledged.

REFERENCES

Cobb, T. (2004). The Complete Lexical Tutor. In TESL – EJ. 8.3. https://www.tesl-ej.org/ej31/m2.html

Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. In Societies. 15 (6). https://doi.org/10.3390/soc15010006

Hakimi, L., Eynon, R., & Murphy, V. A. (2021). The Ethics of Using Digital Trace Data in Education: A Thematic Review of the Research Landscape. Review of Educational Research, 91(5), 671-717. https://doi.org/10.3102/00346543211020116 (Original work published 2021)

Hawkinson, E. (2025). The Automation Abyss. Learning Together.

Hughes, J. (2021). World Link 4. Developing English Fluency. Cengage/National Geographic.

Nguyen Truong Sa, Nguyen Thi Diem Thi, Hoang Ngoc Quynh Nhu, Do Thi Kim Hieu (2025). How AI-Powered Voice Recognition Has Supported Pronunciation Competence among EFL University Learners. Computer-Assisted Language Learning Electronic Journal, 26(3), 64-83. https://doi.org/10.54855/callej.252634

Olsson, E. (2026). Behind the screens: How AI is affecting academic literacy and what we can do about it. Online Presentation for the University of Warwick "Language Culture Matters Seminar Series." February 11, 2026. https://warwick.ac.uk/fac/soc/al/events/seminarseries/

Rodgers, J.C. (2008) Moving Beyond the Plateau: From Intermediate to Advanced Learning.

Cambridge University Press.

Thi-Nhu Ngo, T., Hao-Jan Chen, H., & Kuo-Wei Lai, K. (2024). The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. ReCALL, 36(1), 4–21. doi:10.1017/S0958344023000113

Williams, M., and Burden, R.L. (1997). Psychology for Language Teachers. Cambridge University Press.

Zimmerman, B. (1995). Self-Efficacy and Educational Development. Chapter 7. In Bandura, A. (Ed.) Self-Efficacy in Changing Societies. Cambridge.

APPENDIX

Links to resources mentioned in this paper

Varun Puri at TED X St George *How AI can help you improve your public speaking.*
https://youtu.be/9CVsdayG2kw

Yoodli Presentation Practice
https://yoodli.ai/use-cases/public-speaking

Free Speed Reading Test by AceReader
https://www.freereadingtest.com/

English Vocabulary Size Test: How many words do I know?
https://www.vocabularytester.com/vocabulary-test

Lex Tutor Vocabulary Levels Test (Productive)
https://www.vocabularytester.com/vocabulary-test

Spreeder App
https://www.spreeder.com/app.html