Artículo de Experiencia de Aula
ISSN: 2463-0624 / Vol. 1 Nº 10

Revista Electrónica TicALS

AI-Supported Thinking: Improving Mathematical Reasoning and Communication with Snorkl App

Lina Marcela Gómez Quintero
Magister en Ciencias Naturales y Matemáticas
Universidad Pontificia Bolivariana sede Medellín
lgomez@amigosporvida.com

Fecha de recepción del artículo: (1 marzo 2024); Aceptado: (20 abril 2024)

Abstract

This article's aim is to present the results corresponding to the first stage in the implementation process of thinking-aloud routines towards solving mathematical problems using Snorkl App, an AI-supported educational tool whose beta version is available at the moment of submitting this article. The goal of using this platform is to improve the strategic thinking for problem solving in a group of fifth grade students in a public school of Texas, within the school year and prior to their TELPAS (*Texas English Language Proficiency Assessment System*) and STAAR (*State of Texas Assessments of Academic Readiness*) tests. The implementation, carried out throughout an academic semester, has shown an improvement in the students' skills for problem solving and thinking communication. Scaling up the use of this platform along the whole group level is considered as one of the upcoming steps for this process.

Keywords: artificial intelligence, classroom habits, formative assessment, learning experience, thinking aloud.

Resumen

El objetivo de este artículo es presentar los resultados correspondientes a la primera etapa de la implementación de rutinas de pensamiento en voz alta para la solución de problemas matemáticos empleando Snorkl App, una herramienta educativa soportada en inteligencia artificial que se encuentra en versión beta al momento de la presentación de este artículo. El objetivo de utilizar esta plataforma es mejorar el pensamiento estratégico para la resolución de problemas en un grupo de estudiantes de quinto grado de una escuela pública en Texas, en el marco de año escolar y antes de sus pruebas TELPAS (*acrónimo en inglés para Sistema de Evaluación del Dominio del Idioma Inglés de Texas*) y STAAR (*acrónimo en inglés para Evaluaciones de Preparación Académica del Estado de Texas*). Esta implementación, realizada a lo largo de un semestre académico, ha mostrado a la fecha un mejoramiento en las habilidades de los estudiantes para la solución de problemas y la comunicación de sus procesos de pensamiento. Ampliar el uso de esta plataforma a todo el grado académico es uno de los pasos a seguir en este proceso.

Palabras clave: inteligencia artificial, hábitos de aula, evaluación formativa, experiencia de aprendizaje, pensamiento en voz alta.

Introduction

The term Artificial Intelligence, or AI, is becoming widely known in diverse scenarios, although it is still hard to define it properly (Bostrom, 2006; cited by Luckin and Holmes, 2016); the use and learning of AI has become more relevant in the last 10 years, for it is considered a topic of interest within the government priorities: for example, in 2021, the National AI Initiative Act of 2020 became law, providing guidelines for a coordinate program to accelerate AI research and application; the aim of this law is the trustworthy use of AI to prepare the present and future US workforce in different sectors of economy and society (Niemi, H. 2021). Now, in educational scenarios, the AI has also taken a place, not only as a subject itself (Tsz Kit Ng et al., 2022) but as a support to learn other subjects (Le, N. T et al., 2023).

The most famous name associated with AI is ChatGPT, an example of conversational AI that responds to prompts provided by users with different types of products, such as factual answers, compositions, even multimedia generation, however, the nature of an important amount of data generated by ChatGPT is not reliable (Wong, C. 2023), so its use in developing learning scenarios must be curated by educators to make it authentic and rigorous. An alternative for this interaction with AI in learning environments is the use of AI-supported tools to guide student-centered learning (Le, N. T et al., 2023): these tools are powered by AI, and their settings are provided by educators to ensure that the learning path provided is restrained by specific learning goals.

An example of these tools is Snorkl App, an educational AI-supported tool that involves conversational features and data analysis to provide feedback to the students who provide evidence of their thinking to solve a given problem; the students must explain their reasoning path towards the solution, involving a process known as *thinking aloud* (Güss, C. D. 2018), which helps students visualize their thinking and assess their flaws or inconsistencies. This kind of tool could be implemented in the classroom as further resources to help students developing their communicative skills towards problem analysis and solution, which may have an influence in the development of higher thinking skills by scaffolding the thinking process (Gómez, L. 2018), and offering an option to involve a

variety of skills in the solution of any given problem, which would allow the teachers or mentors to observe individual development and carry out a formative assessment process. Moreover, the consistency in this implementation would lead to the acquisition of solid learning habits in the classroom, or classroom habits (Gómez, L. M. 2019), which could set up learning expectations towards other environments, in and out of the learning community.

A special type of learning community that could implement a solution like this is a public school in the State of Texas, since the state demands the application of specific tests to assess the levels of language proficiency and academic readiness of students: The Texas English Language Proficiency Assessment System (TELPAS), and the State of Texas Assessments of Academic Readiness (STAAR®)

TELPAS tests are officially described by the Texas Education Agency (TEA, 2024):

The Texas English Language Proficiency Assessment System (TELPAS) is an English language proficiency assessment aligned to the Texas English Language Proficiency Standards (ELPS). This assessment is designed to assess the progress that emergent bilingual (EB) students make in learning the English language. TELPAS fulfills ESSA requirements for assessing EB students in kindergarten through grade 12 in four language domains: listening, speaking, reading, and writing.

STAAR tests, on the other hand, are officially defined as follows: [1]

The State of Texas Assessments of Academic Readiness (STAAR®) is a standardized academic achievement test designed to measure the extent to which a student has learned and is able to apply the defined knowledge and skills in the Texas Essential Knowledge and Skills (TEKS) at each tested grade, subject, and course […] STAAR helps to ensure that Texas students are competitive with other students both nationally and internationally. Another important function of STAAR is gauging how well schools and teachers prepare their students academically.

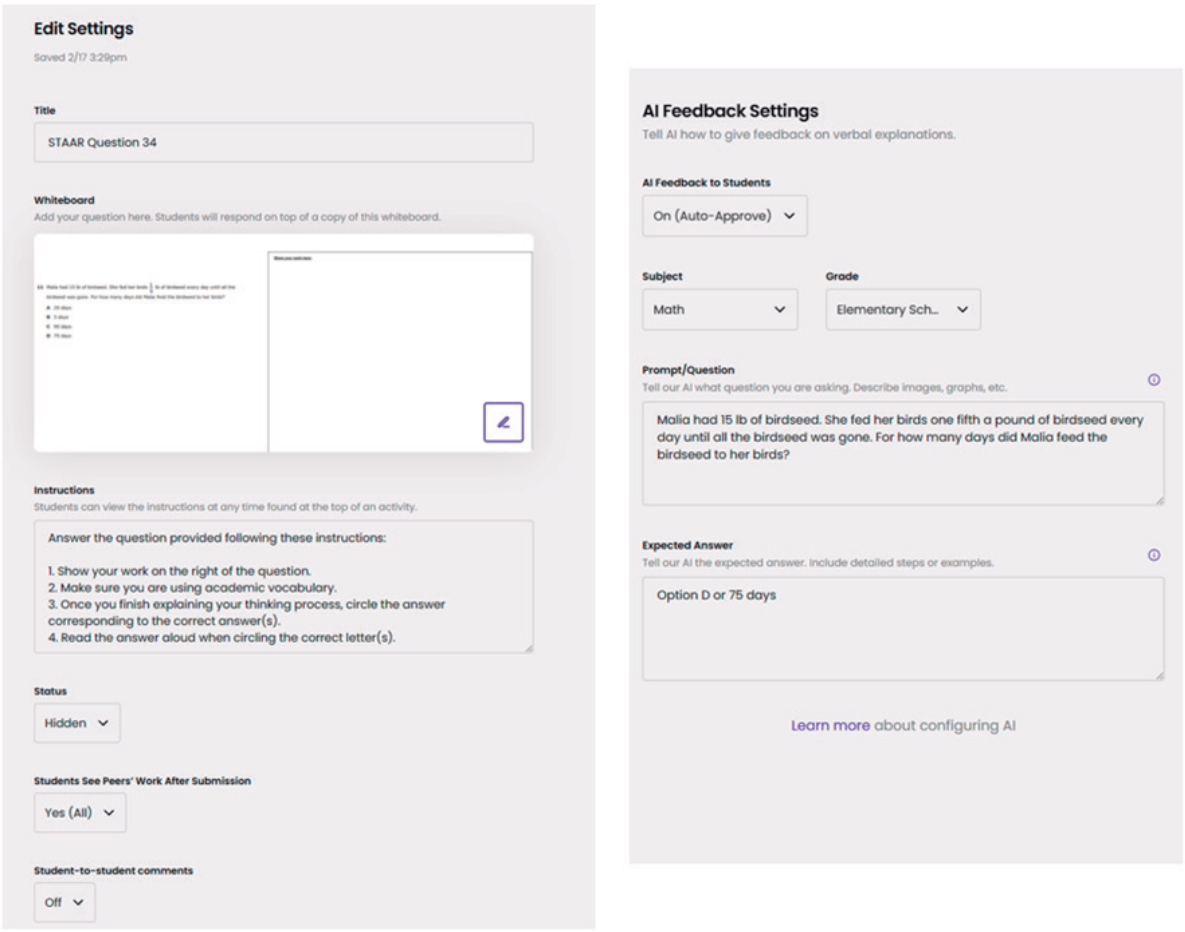
Materials and methods

The hypothesis that led this experience can be stated as follows: if the development of mathematical thinking is strongly connected to the appropriate development of communicate skills in the natural language of the learner (Gómez, L. M. 2018), then providing a learning scenario in which the student uses their natural (or additional) language to approach to given situations that demand their academic skills, receiving immediate feedback to scaffold their practice, will have a positive impact in the development of higher thinking skills, specifically mathematical thinking, which will lead to their readiness for the state assessments. This type of scenario can be provided using an AI-supported solution.

Snorkl App, the platform used for this experience, is in beta version at the moment of submitting this article. It has two interfaces, one for teachers and one for students; the interface for teachers allows the creation of classes and activities by setting up a whiteboard for the students to show their thinking process. For this implementation process, the whiteboard contained a question from the STAAR Released Questions, available in the Texas Assessments Official Site.

An example of one exercise provided is presented in Figure 1.

Figure 1.
Snorkl App, teacher interface to create activities.



Source: own elaboration.

The first stage of this implementation plan has been developed with a sample of 21 students, all of them enrolled in fifth grade. All the students in the sample participated in the following activities:

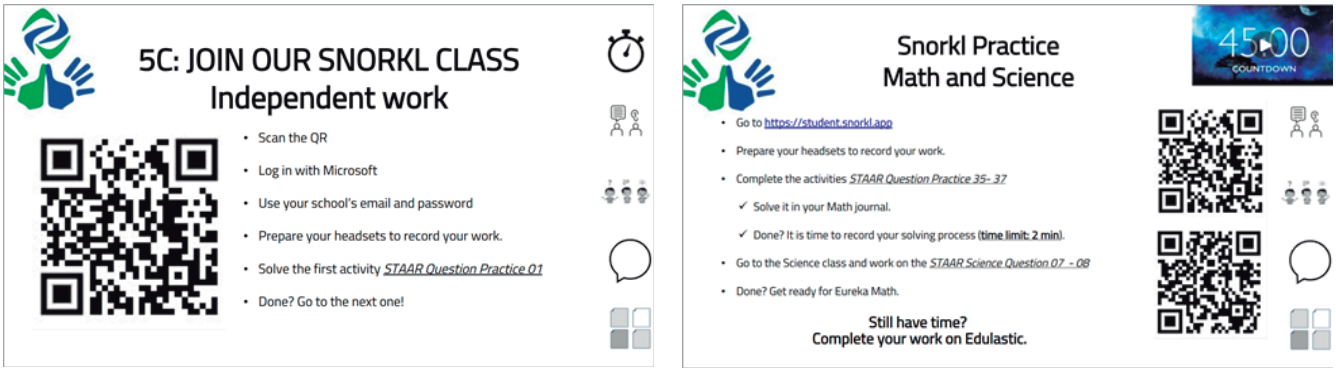
Introduction to the platform: at the beginning of the school year; involved the creation of a sample class for participants to get familiar with the platform.

Practice session protocol: including the instructions to access the platform, assigned practice problems and further directions (Figure 2).

Practice session: twice a week, 45 minutes each session, for six months. Each question or problem assigned should be solved and submitted in the platform, after what the improvement of the think-aloud process using the feedback given by the Snorkl AI had to be executed two more times.[1]

[1] It was possible that the students needed just a fraction of a practice session to solve one or more questions, while others needed more than one week to work on just one question, due to their current communicative skills, as well as to their mastery of specific mathematical skills; that is why several assignments were provided for students to work at their own pace during the assigned practice sessions.

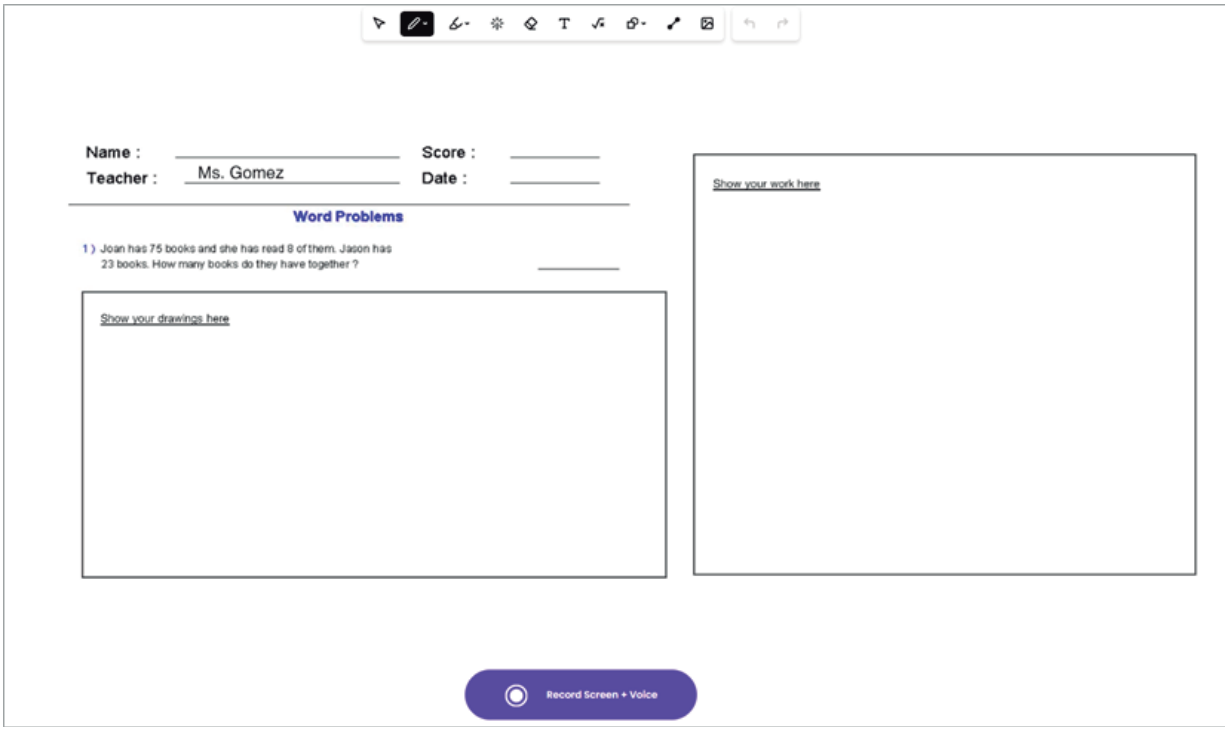
Figure 2.
Snorkl Practice Session Protocol, first and final versions.



Source: own elaboration.

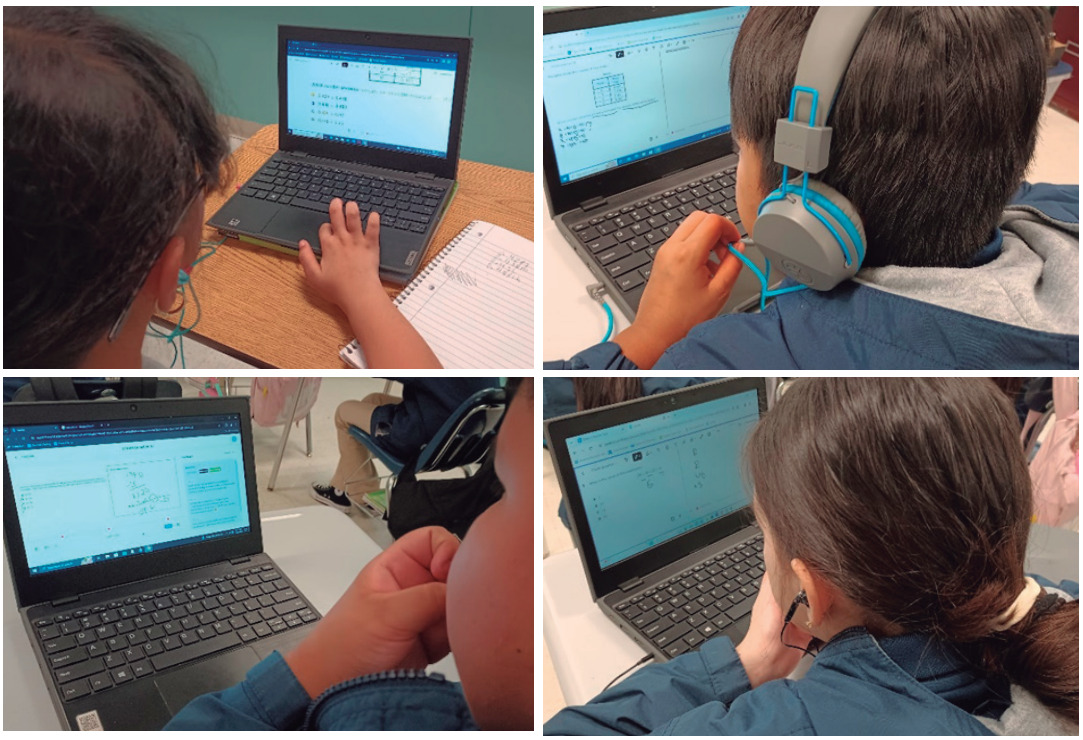
For the development of the practice sessions, each student had a device – tablet or laptop – with internet connection, to access the platform and record their work (Figure 3). Each student interacts with the platform as shown in Figure 4.

Figure 3.
Snorkl App, student interface.



Source: own elaboration.

Figure 4.
Snorkl Practice Session, student work.

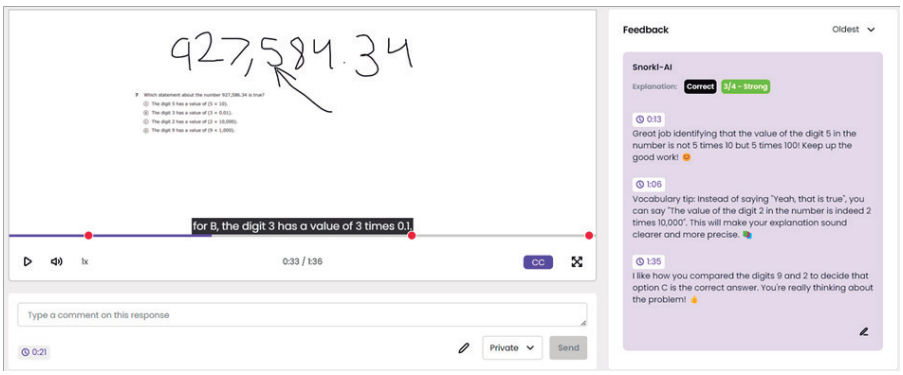


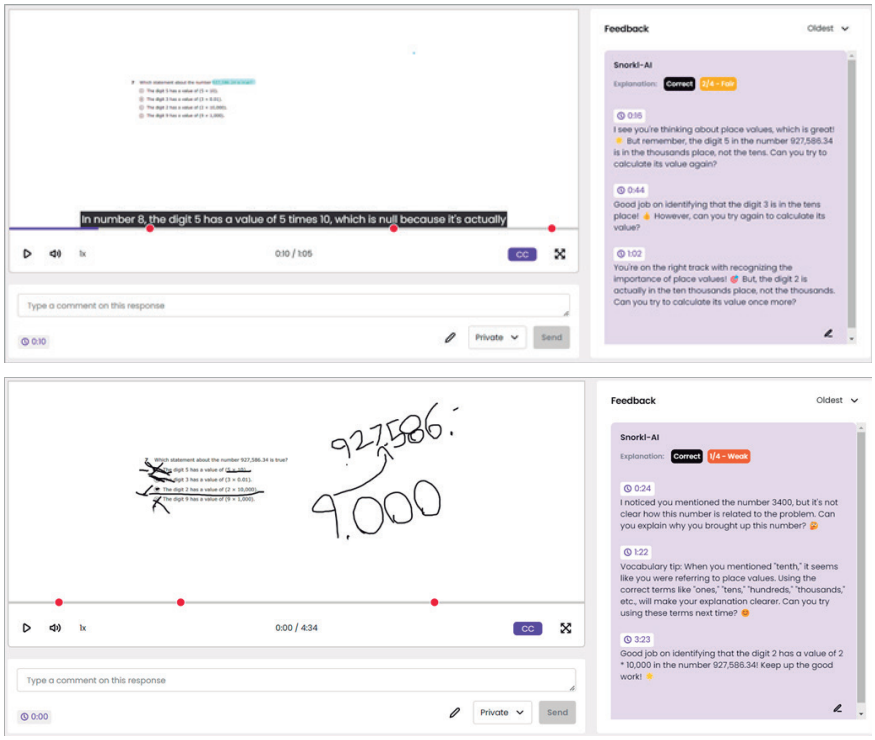
Source: own elaboration.

Results

The platform algorithm that provides assessment, Snorkl AI, gives feedback based on the correctness of the exercise, and the quality of the explanation recorded. Some examples of the feedback given by the app, are available in Figure 5.

Figure 5.
Feedback provided by the Snorkl – AI for three types of submission: strong, fair and weak explanation.





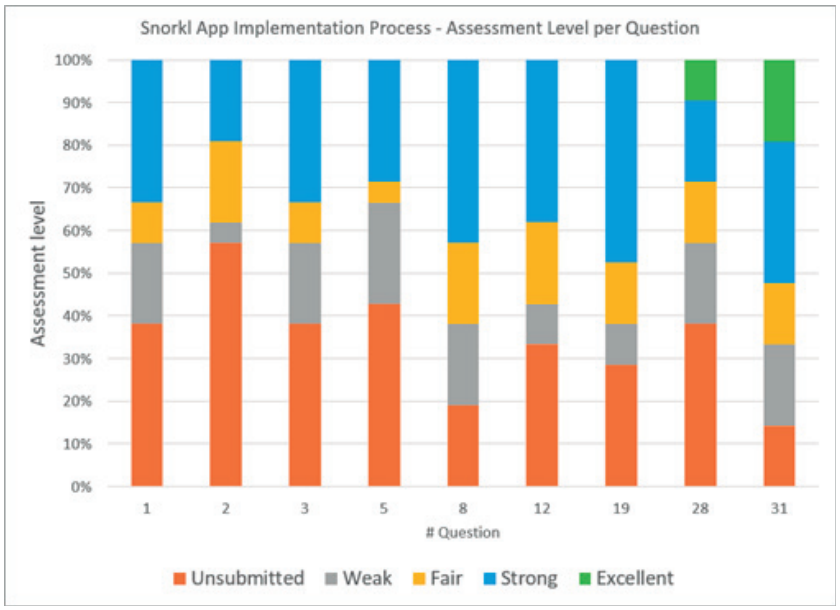
Source: own elaboration.

The conversational features of the Snorkl AI make the feedback accessible to the students, so the highlights about their think-aloud process are easily identified, and the tips or questions to address difficulties are delivered in a simple, effective way. Therefore, it was expected that the consistent exposure to the learning experience with the tool had an influence on the students' habits to solve the questions provided, due to the construction of some elaboration criteria, guided by the feedback provide by the AI.

The changes in assessment levels during different assignments are presented in Figure 6.

Figure 6.

Feedback provided by the Snorkl – AI for three types of submission: strong, fair and weak explanation.



Source: own elaboration.

Results and Findings

It is possible to affirm that the assessment level has improved through time, since the number of submissions evaluated as Fair or Strong has gradually increased through time, while the number of submissions evaluated as Weak has decreased. Moreover, at the end of the recorded timeframe, some submissions have been considered Excellent by the platform, which supports the initial expectation of the progressive building of elaboration criteria with the support of the AI feedback.

In terms of the data analysis for decision making in the structure of lessons or projects, the insights generated by the platform help to identify the needs of the students, towards upcoming lesson planning including spiraled paths, scaffolding, even remediation or whole class intervention. It has been found that the types of insight have varied through time, as shown in Figure 7: while incorrect answers with attempted explanation or no explanation were present at the beginning, they stopped appearing in the report, which suggests an improvement in the abilities of the students to articulate their thinking into words and communicate it verbally.

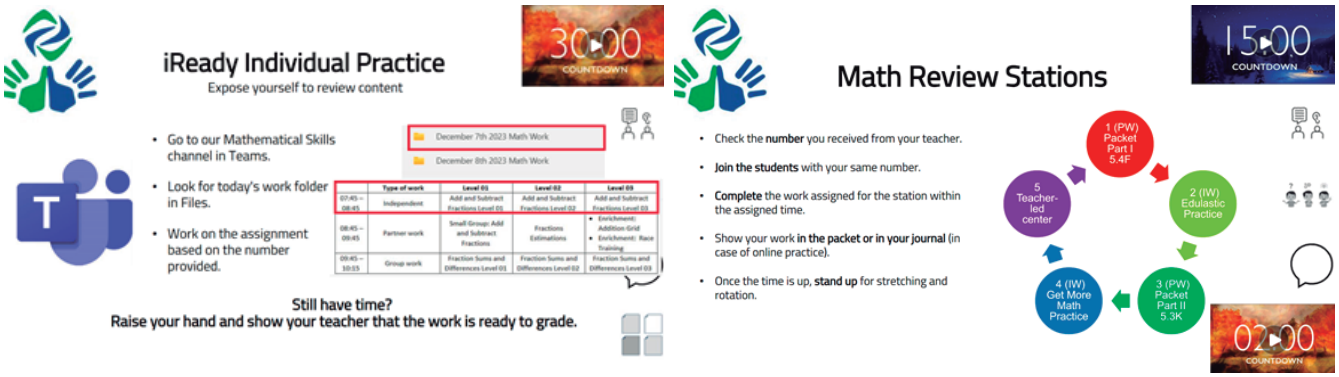
Figure 7.

Example of insights provided by Snorkl App, questions #2 and #34.

Source: own elaboration.

These insights have been discussed with the team, and compared with the data obtained by other sources, such as: tests, problem sets, exit tickets, and others. The consolidation of these data has supported the development of differentiated intervention sessions, with protocols like the ones available in Figure 8.

Figure 8.
Examples of differentiated intervention protocols based on data.



Source: own elaboration.

Discussion and recommendations

During this first stage of implementation, it was found that the participants achieved higher levels of understanding and application of strategic thinking to solve STAAR-related problems or questions, based on the increase of the higher assessment levels obtained during the practice sessions; considering the practice protocols involving the condition of three attempts per exercise, using the questions and tips from the Snorkl AI, it is possible to state that the feedback provided by Snorkl AI helped the students use the improvement opportunities they had when solving each question, which is consistent to the expectations of a formative assessment, building habits to address each assigned task. Besides, it is possible to state that the use of the thinking aloud routines have supported some metacognitive traits in the students, for this observation about their reasoning path through scripting the problem out and explaining it aloud to an algorithm, help finding any inconsistencies or misunderstandings during the activity. This insight leads to the recommendation of scaling up the implementation to the rest of fifth graders in the school.

One week after submitting this article, the participants will take their TELPAS tests, and their STAAR tests will be taken one month later; once the results are released, they will be compared with the previous academic year to confirm the impact of this implementation in the test scores, and then make decisions about scaling up their use in more than one grade level.

Referencias

Gómez, L.M. (2018) *Competencia comunicativa de la matemática: un enfoque para la modelación de situaciones problema.* [Master's degree tesis, Universidad Pontificia Bolivariana] https://repository.upb.edu.co/bitstream/handle/20.500.11912/3928/COMPETENCIA%20COMUNICATIVA%20DE%20LA%20MATEM%C3%81TICA%20UN%20ENFOQUE%20PARA%20LA%20MODELACI%C3%93N%20DE%20SITUACIONES%20PROBLEMA.pdf

Gómez, L. M. (2019) *Educación STEM/STEAM como pretexto para la innovación en comunidades de aprendizaje.* En N. M. Cáceres (Ed.), Educación STEM / STEAM, apuestas hacia la formación, impacto y proyección de seres críticos (p. 56). www.doi.org/10.47212/educacion_stem-steam_5

Güss, C. D. (2018). What Is Going Through Your Mind? Thinking Aloud as a Method in Cross-Cultural Pyschology. Retrieved from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6099082/

Le, N.T, Strickroth, S., Gross, S. & Pinkwart, N. (2023). *A Review of AI-Supported Tutoring Approaches for Learning Programming.* Retrieved from: https://link.springer.com/chapter/10.1007/978-3-319-00293-4_20

Luckin, R., Holmes, W. (2016) *Intelligence unleashed: an argument for AI in education.* Retrieved from: https://discovery.ucl.ac.uk/id/eprint/1475756/1/Pearson%20Intelligence%20Unleash ed%20FINAL.pdf

Niemi, H. (2021). *AI in learning: preparing ground for future learning. Journal of Pacific Rim Psychology, 15 (1).* Retrieved from: https://journals.sagepub.com/doi/pdf/10.1177/18344909211038105

OpenAI (November 30, 2022). *Introducing ChatGPT. OpenAI.* Retrieved from: https://openai.com/blog/chatgpt

Texas Education Agency, TEA (2024). *Student Assessment: testing.* Retrieved from: https://tea.texas.gov/student-assessment/testing

Tsz Kit Ng, D., Lee, M., Yi Tan, R. J., Hu, X., Downie, J. S., Kai Wah & Chu, S. (2022) A *review of AI teaching and learning from 2000 to 2020.* Retrieved from: https://doi.org/10.1007/s10639-022-11491-w

Wong, C. (2023). T*hese scientists aren't using ChatGPT – here's why.* Retrieved from: https://www.nature.com/articles/d41586-023-04071-6#:~:text='It's%20fabricated%20information'&text=%E2%80%9CA%20large%20amount%20of%20information,fabricated%20information%2C%E2%80%9D%20says%20Vittaladevaram.