EDUCATE

VENTURES RESEARCH

# Case Study

# Third Space Learning

# High-impact AI Maths Tutoring

# Executive Summary

This case study provides an independent evaluation of Skye, Third Space Learning's high-impact AI maths tutor. Third Space Learning (TSL) works with UK primary and secondary schools and US elementary and middle schools, providing annual licenses that allow schools to offer every child spoken, high-impact AI tutoring programmes with Skye, their AI tutor.

Following a rapid evidence review, we explored the extent to which Skye is aligned with core evidence principles for high-impact tutoring. In addition, we examined the initial evidence of the impact of Skye on teaching and learning outcomes, and highlighted the evidence of impact, the potential implementation challenges, as well as further opportunities and best practices.

Skye demonstrates strong alignment with core research principles for effective, high-impact AI tutoring. Its "I do, we do, you do" scaffolded structure aligns with step-based learning, breaking complex problems into manageable components. It delivers immediate, targeted feedback through probes, hints, and correction, prioritises guidance over answer generation, and ensures teachers maintain agency over topic selection.

Our evaluation reveals promising initial outcomes across multiple dimensions. Students demonstrated substantial within-session learning gains, improving from 34% accuracy on diagnostic check-in questions to 92% on check-out assessments. Confidence metrics showed similar positive trajectories, with 59.4% of students ending sessions with higher confidence levels and 66% showing session-on-session confidence increases over time. Two learner groups emerged as particular beneficiaries: anxious or quiet pupils who hesitate to seek help in traditional classroom settings found the non-judgmental AI interaction more accessible, and students requiring immediate consolidation after teaching benefited from the targeted, adaptive practice.

Skye demonstrates strong alignment with ethical AI tutoring principles through its curriculum-appropriate content, user agency preservation, and age-appropriate dialogue. The system maintains human-centered learning with teachers retaining professional judgment over topic selection and student participation. However, challenges remain, particularly with regards to inclusion such as in the case of Additional Language (EAL) learners and students with speech recognition difficulties who sometimes find it difficult to communicate with Skye.

# Research Aims and Approach

This evaluation seeks to understand the effectiveness and implementation of Skye through three primary research questions. First, we examine the implications of the emerging evidence base for AI tutoring effectiveness, including how tutoring can support various learning outcomes and the key ethical and safety considerations. Second, we assess how Skye's product design aligns with research-led guidance on safe, ethical and effective high-impact AI tutoring. Third, we investigate the perceived impact of Skye on various learning outcomes, specifically considering affect, skills, and knowledge acquisition.

# Methodology

Our approach is evidence-led, beginning with a rapid review of the literature on high-impact AI tutoring effectiveness and one-to-one mathematics pedagogy. This evidence base informs our research questions and data collection strategy, providing a framework against which to evaluate Skye's design and impact.

To evaluate Skye, we collected data from teacher and student experiences, as well as how students actually worked with Skye. Our mixed-methods approach combined quantitative analysis of tutoring sessions with qualitative insights from focus groups and surveys. In total we had:

1. Platform analytics from thousands of sessions, including entry/exit checks, tool use (hints, talk-moves, corrections), and confidence prompts.
2. Teacher interviews across 6 primaries with varied cohorts and implementation models (in-class, clubs, catch-up blocks).
3. A short student survey on usefulness, enjoyment, understanding, independence, exam readiness, and maths anxiety.

Across schools the implementations often differed, dedicated computer rooms vs laptop trolleys, tight supervision vs light-touch oversight.

# Alignment of Skye with core research principles on effective high-impact tutoring

## a) Skye use-case

Skye is an AI maths tutor developed by Third Space Learning (TSL), providing spoken, one-to-one high-impact Maths tutoring programmes, reflecting the research and best practices of high-impact tutoring. Students interact with Skye via a shared screen and microphone; they speak and listen as Skye guides them through problems. Lessons follow an “I do, we do, you do” scaffolded structure. They are created by UK and US teachers, and cover UK national curriculum at Key Stage 2 and 4 (with Key Stage 3 to follow), as well as US Common Core (state-aligned) in Grades 3-8 (Elementary and Middle).

Skye sessions start with a diagnostic, a set of 'check-in' questions about a particular maths topic which allows the AI tutor to then tailor the rest of the session to the student's understanding. Teachers can also reorder lesson topics and select programmes to align with classroom goals. Skye's tone is warm, optimistic and encouraging. Teachers also describe Skye's dialogue as patient and structured. The conversation proceeds through a set of 'slides', with Skye introducing each slide and explaining the topic. Students can interrupt Skye at any time, and it addresses their questions immediately. This is then

followed by a set of questions to test the student's understanding of the topic. Pupils usually enter their answers on the screen, but also speak answers aloud or ask further questions if they are confused about something; Skye probes, hints, and corrects before moving on.

Skye's highly moderated AI can only access lesson content created by Third Space Learning. Data collected by the tool is not used to train the model, and only authorised staff can schedule sessions.

## b) Evidence-led product design

The rise of Artificial Intelligence (AI) in education marks a transformative era, offering the promise of personalised and adaptive learning experiences. With significant energy now being invested into developing and using AI tutors, the challenge for educators and developers is to harness their full potential in ways that truly benefit learners. Decades of research on human-led, high-impact tutoring consistently demonstrate its strong impact on learning outcomes (Bloom, 1984), and an established evidence base for Intelligent Tutoring Systems (ITS) shows that well-designed systems can deliver comparable gains in structured domains (Anderson et al., 1994; du Boulay, 2016).

The research highlights several core principles for learning design of high-impact AI tutors. One is the power of step-based learning: breaking problems into manageable steps accelerates progress and helps students achieve mastery far more efficiently than traditional instruction (Anderson et al., 1994). Another is the importance of immediate, targeted feedback—learners benefit most when feedback is concise, specific, and focused on process rather than lengthy explanations or simulated conversation (du Boulay, 2016). A third is support for multiple solution paths, recognising that students may approach problems in different, but equally valid, ways. AI tutors should focus on guiding learners through steps, offering process-focused feedback, and adapting flexibly to diverse problem-solving strategies.

There are multiple implications for the learning design of AI tutors. AI tutors should be designed around process guidance (high-impact tutoring principles) rather than answer generation (homework help or AI learning assistants), with feedback that is timely, specific, and concise. Implementation strategies must prioritise teacher training and adaptation, positioning AI as an assistant rather than a substitute. Finally, evaluation should go beyond test scores to include measures of time to mastery, motivation, and the transfer of skills beyond the AI environment.

Implementation studies also provide key insights. The most successful uses of ITS occurred when they were integrated into classrooms as complements to teaching, not replacements (du Boulay, 2016). Teachers play a vital role in shaping how AI is used, but evidence shows they often need one to two years to adapt and fully embed these tools into practice (Anderson et al., 1994). Blended learning models—where AI supports foundational skills and frees teachers to focus on higher-order tasks—have consistently

shown the strongest impact. Just as importantly, student engagement has proven significant: when learners see steady progress, they are more motivated, and struggling students in particular show substantial gains (du Boulay, 2016).

Through its “I do, we do, you do” scaffolding structure, Skye demonstrates a step-based learning model on breaking a problem into manageable steps. Additionally, Skye’s lessons show a progression from demonstration to guided practice to independent work that has been shown to be effective in ITS research. The diagnostic check-ins at the start of the lesson enhance this further by allowing Skye to calibrate itself for each student depending on their capability.

The tool’s approach to immediate, targeted feedback also aligns with research recommendations. Rather than providing lengthy explanations, Skye uses probes, hints, and corrections to guide students through problem-solving processes. The voice-first interaction model also makes it so the feedback is usually concise and focused.

Skye's deployment model follows the research consensus that AI tutors work best as complements to human teaching. The system preserves teacher agency through allowing educators to select topics, reorder lessons, and determine which students participate. This positions Skye as what one teacher called a "force-multiplier for practice".

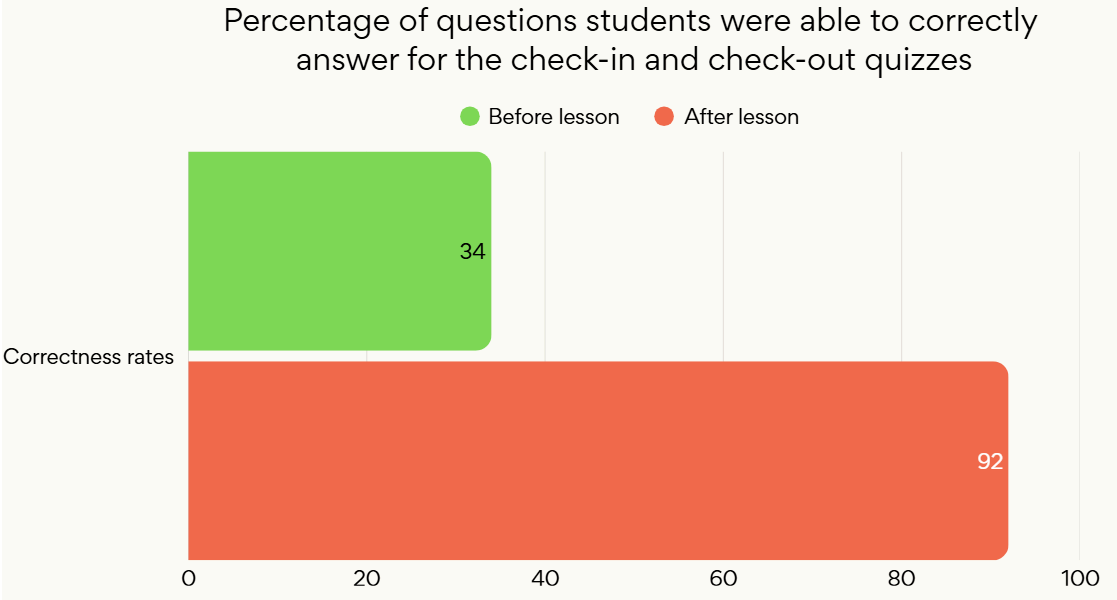
# Initial Evidence of Impact

The following section outlines the findings from the data analysis, which encompassed data from 9,320 student sessions with the AI tutor, interviews with 6 teachers on the use of Skye with their pupils, and a survey of 14 students on their experience with Skye.

## a) Learning Outcomes

### i) Attainment

Across the 9,320 sessions with the AI tutor, the 'check-in' or diagnostic questions that happen at the start, pupils were only able to answer 34% of the questions correctly. However, the same students were able to answer 92% of the questions (judging their understanding of the topic) at the check-out questionnaire; consistent with meaningful within-session learning and consolidation.

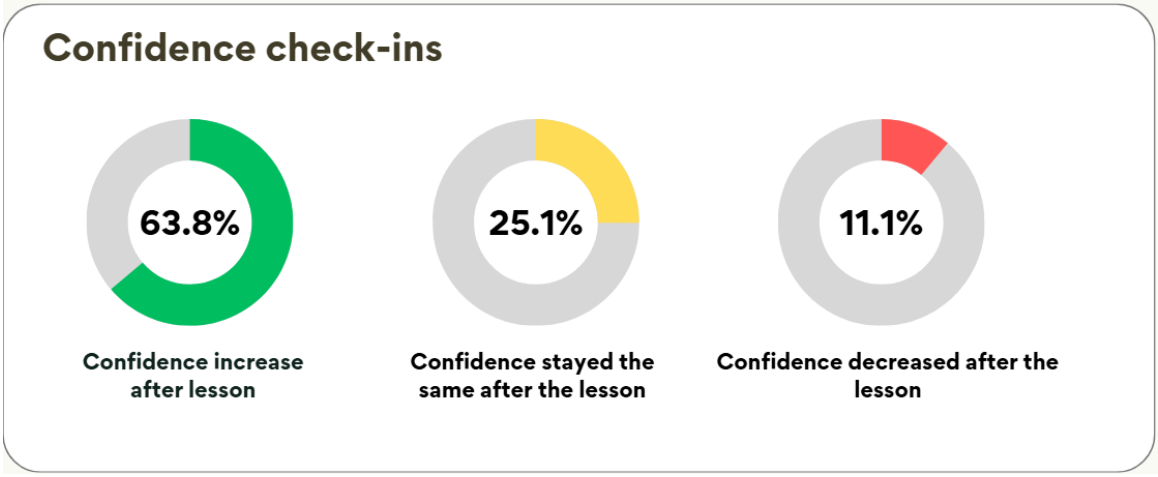



Check-in → check-out gains within sessions are large; confidence typically rises by session end and trends upward over repeated use. Pupils rate Skye as moderately useful, enjoyable and clarity-enhancing, and these perceptions move together. Heavy use of hints, talk-moves - structured prompts to encourage pupils to explain their reasoning - and corrective feedback suggests pupils are not just hearing answers but being guided through steps and strategies.

The pattern is consistent with guided practice that builds fluency and self-efficacy. We are not yet estimating standardised attainment impact; that requires a controlled design. But the triangulation, check-out accuracy, confidence shifts, perceived usefulness/understanding, and teacher observations, points to real, near-term learning value inside sessions.

### ii) Motivation and Confidence

Confidence amongst students tended to rise. In sessions with confidence prompts, 63.8% of students ended higher than they began, 25.1% were unchanged, and only 11.1% students showing a decrease in confidence. Tracking the same pupils over time showed a similar pattern: roughly 66% showed an increase in confidence session-on session, 26% held steady, and 9% decreased across multiple sessions.

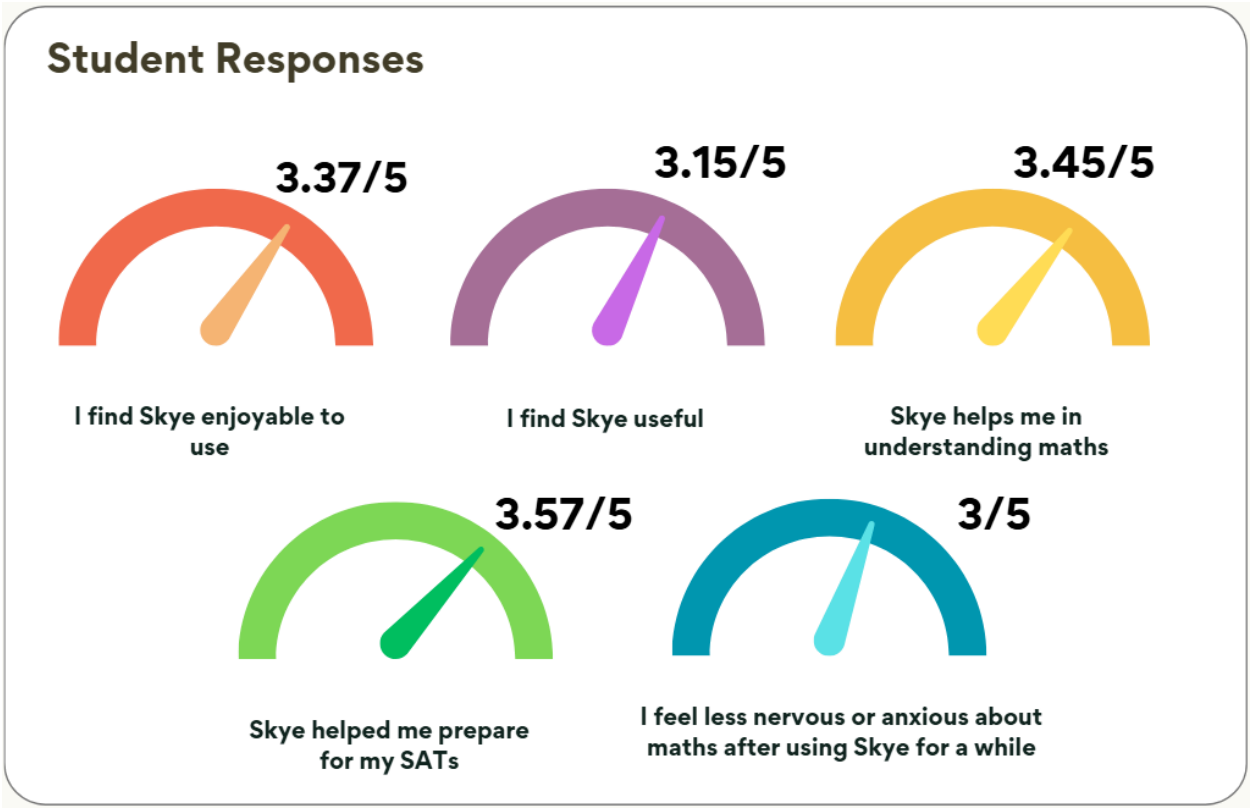



Leaders describe Skye as a "launchpad" for pre-teaching and just-taught consolidation; several noted that pupils using Skye "performed better than those that haven't," not as a causal claim, but as a practical observation tied to confidence and willingness to attempt problems.

**iii) Affect**

Survey responses were moderately positive: students found the tool useful (mean rating 3.4 out of 5), enjoyable to use (3.2), that it helped their understanding of maths (3.5), and easy to use independently (3.6). Students also reported a preference for talk-and-listen learning (3.6) and comfort with technology for learning (3.7). On exam-readiness, the students felt that it helped them prepare for their SATs (3.6). It also slightly helped students with their nervousness (3.0).

Teachers also emphasised that pupils who are anxious about being wrong in front of peers are "more willing to talk to an AI," which in turn unlocks participation and practice.




#### iv) For Specific Sub-groups of Learners

Two groups stood out. First, quiet or anxious pupils who hesitate to ask for help in class. One leader noticed they were "more willing to engage with an AI tutor... more willing to take a risk" because a computer's correction felt less judgmental than a human's. Second, pupils who benefit from immediate consolidation after teaching. With strand-level targeting or diagnostics, Skye became a launchpad for pre-teaching and just-taught practice. At Danegrove, once routines settled, roughly two-thirds of a small pilot group "liked it, enjoyed it, looked forward to it."

Pupils with attention/behaviour needs and those with limited English proficiency require closer adult presence and targeted scaffolds. Device/microphone quality and room acoustics are equity issues: without quiet spaces and headsets, speech-recognition errors rise and learning stalls. Equity here is infrastructural (space, kit) and pedagogical (routines, language scaffolds).

## b) Teaching Outcomes

#### i) Personalisation

One of the main advantages of Skye was reported to be its ability to tailor its learning to the context and ability of the learner. In fact, in schools that had run

their own evaluations of Skye, it was reported that students that usually struggled in regular maths lessons, were the ones to benefit most from the personalised help that Skye was able to offer.

**ii) Availability**

For teachers, the biggest advantage of the AI tutor was availability, affordability, and patience. Skye delivers a stable lesson arc, repeats without judgement, and makes time-on-task easier to schedule.

# Ethics and Safety

In terms of ethics and safety, we analysed Skye against several ethical dimensions with the aim of establishing whether it maintained human-centered learning, protected student well-being, ensured fairness and inclusion, and had clear governance and accountability structures in place.

## a) Learning outcome alignment

With its diagnostic check-in and check-out questions, confidence check-ins, and heavy scaffolding, Skye is able to adapt to individual student needs and so supports diverse learning objectives beyond simple correctness.

## b) User Agency Preservation

Skye's design allows students to interrupt at any time and ask questions, maintaining learner control over the pace and direction of learning. Teachers retain professional judgment in selecting topics and determining which students participate. As one teacher noted, Skye delivers "a stable lesson arc" but teachers maintain ultimate decision-making authority.

## c) Emotional Intelligence and Well-being

Several teachers praised Skye for being able to serve more anxious students well, with teachers reporting students were "more willing to talk to an AI" because "a computer's correction felt less judgmental than a human's." The tool is able to detect when a student displays anxiety or stress, and has a patient, encouraging tone when teaching students.

## d) Age-appropriate and Safe Implementation

Skye uses curriculum-aligned content appropriate for year groups, employs the Concrete-Pictorial-Abstract (CPA) approach suitable for the developmental stage, and maintains warm, optimistic, encouraging dialogue calibrated for primary students. It operates within strict guardrails, and is able to steer the conversation back to the topic if the lesson is being diverted.

However, there are also certain areas where improvements could be made. Specifically:

### a) Tailoring to EAL learners

In teacher interviews it was suggested that Skye has challenges with EAL (English as Additional Language) students. Teachers report "children that don't speak English particularly well...really struggle" and that the system has difficulty with "very quiet speakers, children in noisy rooms, and early-stage EAL learners."

### b) Adaptive Human Interaction Balance

TSL's high-impact AI tutoring programmes are delivered in school as a part of the school timetable, reflecting the research and best practices in high-impact tutoring. All the sessions we analysed occurred in controlled settings in schools, and most teachers we spoke to emphasised the benefit of human oversight. This was especially noted with students with disruptive behaviours who would easily get distracted and try to divert the AI tutor from the lesson.

## Implementation Considerations

Almost all teachers emphasised the importance of the role of the class teacher as the human in the loop to oversee goal-setting, behaviour, motivation and safeguarding. This reflects a fundamental design principle for TSL's high-impact AI tutor programmes, with the tutoring programmes directed by teachers to reinforce class teaching, as a core intervention strategy in the school.

This is particularly important where implementation challenges occur. One senior leader stated: "You can hear Skye clearly, but it doesn't always hear the children. That's when an adult keeps things moving." In several settings, behaviour was the limiting factor: a coordinator noted that some pupils "were just trying to get the computer to do silly things," a pattern amplified when learners knew there wasn't a human on the other end. Skye was best deployed as a force-multiplier for practice, not a replacement for expert teaching; no setting treated Skye as a fully autonomous solution.

The voice model is easy for pupils to understand. One headteacher contrasted it with human tutoring (a precursor to AI tutoring, provided on the same platform): "Skye is using a very much a non-accented voice, and the children find that a lot easier to understand"; the harder challenge is Skye understanding pupils, especially very quiet speakers, children in noisy rooms, and early-stage EAL learners. Skye doesn't "read the room." Speech recognition remains a practical constraint in noisy rooms and with very quiet speakers; one teacher described Skye mishearing "twenty-five" and marking it wrong. One teacher trained pupils to "just try again" rather than over-enunciate when the system asked for a repeat. Practical mitigations include good headsets, quiet spaces, sentence stems, and repeat-backs ("I heard twenty-five, correct?").

Early pilots also surfaced latency and pacing issues, "that buffering thing," and moments where progress felt too slow or too fast. A deputy head who listened in on sessions praised the curricular alignment and steady improvement but flagged latency as the biggest risk to focus: "Kids will wait for only so long---if it spins, you lose them."

One teacher recalled early weeks where "the system freezes... Skye doesn't hear them... that buffering thing," But emphasized that this was followed by noticeable improvements to Skye as noise handling and speed got better.

## Conclusion

Our evaluation sought to measure the impact of Skye across three primary research questions: the implications of the emerging evidence base for high-impact AI tutoring, the alignment of the design and implementation of Skye with research principles, and the perceived impact on learning outcomes for students using Skye.

The emerging evidence-base points to the potential of high-impact tutoring to improve learning outcomes when designed in alignment with established principles of learning design. Our evaluation of Skye indicates that the learning design principles have been carefully integrated into the AI Tutor. For example, Skye relies on an "I do, we do, you do" scaffolded structure and emphasises step-based learning and immediate, targeted feedback.

The data collected indicates improvements in student learning outcomes, both in terms of learning gains within sessions, but also in improved confidence over time. In particular, Skye has proven to be valuable for students who may have traditionally struggled to access support: for example, anxious or quiet pupils who hesitate to seek help in classroom settings have benefitted from the non-judgemental nature of Skye, enabling them to engage more fully with mathematical concepts.

Teachers have reported that Skye offers consistent availability and personalised learning, and has been effective in improving the performance of students who are struggling in class. At the same time, there remain some areas of improvement, particularly in terms of accessibility and usability (for example, for EAL learners). Additionally, the most effective interventions were those that ensured appropriate infrastructure such as quiet spaces and quality headsets.