AI Friction Labs

AI FRICTION LABS · CASE STUDY · SPRING 2026

What a Test Can't See

Friction Bot: The Lost Tourist · Pinkerton Academy · Spanish 1 & 2

The Setup

Emily Caseres teaches Spanish 1 and 2 at Pinkerton Academy in New Hampshire. In April 2026, she ran the Lost Tourist Friction Bot — a structured interaction that places students in real-world Spanish scenarios and refuses to speak English — with two classes: a Spanish 1B group of roughly eleven students and a Spanish 2A group of nineteen. Students chose their own scenario from a menu that included ordering at a restaurant, navigating to a destination, and describing medical symptoms. The transcript of each interaction served as the assessment.

Caseres went in skeptical. She had just covered food vocabulary with her Spanish 1 class — timing that made the restaurant scenario a natural fit — but she wasn't sure how either group would respond. "I was a little bit skeptical after my Spanish 1 class used it," she said afterward. "But I was kind of surprised when I got to my Spanish 2 class."

What the Transcripts Showed

Across twenty sessions, the strongest and most consistent finding was this: persistence outperformed accuracy. Students with significant grammar errors — wrong conjugations, phonetic approximations, broken syntax — scored as high or higher than students with cleaner Spanish who kept their responses short and safe.

Brody Yennaco, a Spanish 1 student, typed 'aqua' for agua and 'apetizir' for appetizer throughout his session. His grammar was consistently poor. He navigated a genuine misunderstanding about the type of restaurant, ordered a multi-course meal, and corrected a confusion about his order mid-conversation. He scored 91/100. James Demers, in the Spanish 2A class, scored only 24 out of 40 on Language Proficiency. He scored 38 out of 40 on Persistence and Adaptability, and 30 out of 30 on Goal Achievement. His grammar did not stop him from getting what he needed.

> *"Even when I didn't know specific words, I found ways to keep the conversation going. I made a few mistakes and hesitated at times, but overall I was still able to complete the task."*
> — Spanish 2A student

This pattern repeated across the dataset. The Friction Bot was not measuring grammatical correctness — it was measuring functional communication under pressure. Whether a student could adapt when a conversation stalled, recover from a misunderstanding, and keep moving toward a goal in the target language. A written test shows whether a student can conjugate correctly in isolation. The transcript shows whether they can function when it matters.

The Transcript as Record

David Landaverde's session on April 17 illustrates what the Friction Bot captures that a test cannot. He opened a session in French, responded in Spanish, produced eight low-effort messages, and scored 15/100. Within the same class period, he opened a new session in Spanish, ran for seventy-four messages, and scored 85/100. The bot did not penalize him for the first attempt. It waited. His extended session moved through food, cars, and football, navigated multiple misunderstandings, and showed what happens when a student commits to improvising with limited vocabulary. The arc from abandoned attempt to sustained engagement — visible only because the full transcript exists — is the kind of evidence a quiz cannot produce.

Three students' transcripts also flagged specific messages as containing pasted text — a mid-conversation translator assist, an opening prepared sentence, a phrase inserted during a difficult exchange. Each one marks an exact moment: the student hit a ceiling, and the transcript records what they did next. That is information a test answer sheet does not provide.

AI Friction Labs

Two Classes, Two Outcomes

The experience was not uniform across both classes, and the difference matters for how the Friction Bot is deployed. In Spanish 2A, fifteen of nineteen students rated their performance as good. The average usefulness rating was approximately 3.7 out of 5. Twelve of nineteen said the activity was more realistic than a typical in-class activity. Students described staying in Spanish beyond what the task required and developing their own language strategies.

79%	3.7/5	63%	2.5/5
Spanish 2A rated performance "Good"	Avg. usefulnessSpanish 2A	Spanish 1B rated performance "Weak"	Avg. usefulnessSpanish 1B

In Spanish 1B, the dominant complaint was not that students struggled to produce Spanish. It was that they could not understand what the bot was saying. "Because I couldn't understand the bot," wrote one student. "I only understood a little bit of what the robot was saying and didn't know how to reply," wrote another. Productive friction requires comprehension before it can require adaptation. When a student cannot parse the input, the loop breaks.

The general tourist bot was not calibrated to novice or intermediate-low proficiency — its register was too high for where Spanish 1B students actually were. Caseres recognized this independently, noting that she could imagine differentiated versions built for each level. The 1B data is what a calibration mismatch looks like in practice. It is a deployment lesson, not a verdict on the product.

What the Teacher Found

Caseres used the AI-generated assessment summary to cross-check her own observations — not to replace them. "I really liked the summary of the conversation, because it gave me a second perspective — let me see if I'm noticing the same things." She was not looking for the bot to grade her students. She was looking for a record she could compare against her own read.

> *"It helps me to see more their proficiency versus their performance — because I, as the teacher, know exactly what we've been working on. It was helpful to have something that gives them that space where they have to work with the language they have in a more authentic setting."* — Emily Caseres, Spanish Teacher, Pinkerton Academy

She also named a gap that the Friction Bot addresses that she cannot fill herself. In a classroom, students know the teacher. They know, eventually, that she understands what they are trying to say. The bot does not. It holds the position of a sympathetic listener who is genuinely waiting for communication to happen.

> *"I don't have as many opportunities to help them go and try to speak to another sympathetic listener. They know it's me — they know eventually I actually know what they're saying."* — Emily Caseres

Making Thinking Visible

A language test measures whether a student can produce correct forms under controlled conditions. The Lost Tourist Friction Bot measures whether a student can function in the language when the conditions are not controlled — when the conversation goes somewhere unexpected, when the vocabulary runs out, when the other person keeps talking and does not slow down.

The strongest sessions in this dataset were not the ones with the cleanest Spanish. They were the ones where students simplified their sentences, invented approximations, circled back to clarify, and kept going. One Spanish 2A student described it this way: "I had to take a second to think about the words and connect them together with other words I knew. I had simplified the sentences down." That is circumlocution. That is a real language skill. And it showed up in a transcript, not on a quiz.

Students do the thinking. The Friction Bot provides the resistance. The transcript becomes the evidence.