Exploring the Potential of Speech Recognition to Support Problem Solving and Reflection

Wizards go to school in the elementary maths classroom

Manolis Mavrikis[1], Beate Grawemeyer[2],
Alice Hansen[1], and Sergio Gutierrez-Santos[2]

[1] London Knowledge Lab, Institute of Education, University of London, UK
`{m.mavrikis,a.hansen}@ioe.ac.uk`
[2] London Knowledge Lab, Birkbeck College, University of London, UK
`{beate,sergut}@dcs.bbk.ac.uk`

Abstract. The work described in this paper investigates the potential of Automatic Speech Recognition (ASR) to support young children's exploration and reflection as they are working with interactive learning environments. We describe a unique ecologically valid Wizard-of-Oz (WoZ) study in a classroom equipped with computers, two of which were set up to allow human facilitators (wizards) to listen to students thinking-aloud while having access to their interaction with the environment. The wizards provided support using a script and following an iterative methodology that limited on purpose their communication capacity in order to simulate the actual system. Our results indicate that the feedback received from the wizards did serve its function i.e. it helped modify students' behaviour in that they did think-aloud significantly more than in past interactions and rephrased their language to employ mathematical terminology. Additional results from student perception questionnaires show that overall students find the system suggestions helpful, not repetitive and understandable. Most also enjoy thinking aloud to the computer but, as expected, some find the feedback cognitively overloading, indicating that more work is needed on how to design the interaction tipping the balance towards facilitating post-task reflection.

Keywords: Speech Recognition, Wizard-of-Oz, Intelligent Support

1 Introduction

The importance of language as both a psychological and cultural tool that mediates learning has been long identified; from as early as Vygotsky to modern linguists such as Pinker. From a Human Computer Interaction (HCI) perspective, speech recognition technology has the potential to enable more intuitive interaction with a system, particularly for young learners who reportedly talk aloud while engaged in problem solving (e.g. [1]). However, with the exception of limited research discussed in Section 2, the relationship between speech and learning has not been investigated in the Technology-Enhanced Learning (TEL) field, despite the advances that automatic speech recognition (ASR) is making.

The potential of such technology for learning relates to young children's capacity of inner-speech (that can become explicit think aloud) and reflection as a learning mechanism. For example, in mathematics, our area of interest, several researchers have emphasised that metacognitive instruction that uses self-directed speech improves students' mathematical reasoning [2]. Elementary school students can be aware of their inner-speech and verbalise their thoughts [1]. While the importance of explicit reflection is well understood [3], researchers have also suggested that reflecting is the answer to stimulating retention and that the skill of reflection must be taught at an early age [4].

In this paper we report studies performed in order to understand better the potential of speech-enhanced intelligent support in young children's interactive learning in the iTalk2Learn system, a platform that sequences tasks from intelligent tutoring and exploratory learning environments, while leveraging speech recognition technology to enhance the intelligent decision making, and speech production to provide feedback for students.

The main research question that concerns us: will students talk naturally to the computer and think-aloud, when seemingly unobserved? In the long term, we are interested in the potential of employing speech-recognition to enhance the system's ability to support student problem-solving, exploration and reflection.

The main research question that concerns us is whether providing speech-enhanced intelligent support both at cognitive and affective aspects can demonstrate potential for supporting student problem-solving, exploration and reflection. Additionally we were interested in whether students would talk naturally to the computer and think-aloud, when seemingly unobserved.

We designed and conducted a Wizard-of-Oz (WOZ) study where wizards simulated realistic capabilities of speech recognition technology that shows promising results with respect to both encouraging students' verbal reflections and providing problem-solving support. To the best of our knowledge, the study was unique in its complexity and ecological validity in that it was conducted in a school classroom targeting in six sessions two students at a time while the rest of the classroom was working on the same system (with limited support) thus allowing the students to feel unobserved and think aloud and talk to the computer as they would normally do in the system proper.

In what follows, Section 2 presents a brief background with respect to the potential of voice interaction for learning. Section 3 outlines the methodology and the setting behind the WOZ studies. Section 4 presents descriptive results and an exploratory analysis of the effect of feedback on students' reactions. Section 5 discusses the results in more detail and Section 6 concludes the paper and draws implications for future research.

2 Background — Voice Interaction For Learning

From an HCI perspective speech production and recognition can provide potentially more intuitive interaction. In particular, spoken language input can enable students to communicate verbally with an educational application and thus interact without using human interface devices such as a mouse or keyboard.

Despite ASR for children being extremely difficult it is worth bearing in mind that related HCI-research suggests that ASR accuracy should not limit its usage and that the overall VUI design and the match of the application to its context should be able to compensate for possible flaws [5]. The approach taken in previous work (particularly the LISTEN [6] project) suggests that 100% accuracy can neither be expected nor relied upon. In light of this, it is preferential to err on the side of caution, thereby ensuring the least negative impact on learning.

With respect to learning in particular, the hypothesis that ASR can facilitate learning is based mostly on educational research that has shown benefits of verbalization for learning (e.g., [7,8,9]). The possible verbalization effect could be enhanced with ASR since cognitive theory of multimedia learning [10] predicts that a more natural and efficient form of communication will also have positive learning gains. The few existing research studies have found mixed results with respect to whether the input modality (speaking vs. typing) has a positive, negative or no effect on learning. In [11], for example, the authors investigated whether student typing or speaking leads to higher computer literacy with the use of AutoTutor. They reported mixed results that highlight individual differences among students and a relationship to personal preferences and motivation.

The importance of students' verbal communication in mathematics in particular becomes apparent if we consider that learning mathematics is often like learning a foreign language. Focusing, for example, on learning mathematical vocabulary, [8] encouraged students to talk to a partner about a mathematical text to share confusions and difficulties, make connections, put text into their own words and generate hypotheses. This way, students were able to make their tentative thinking public and continually revise their interpretations.

For further consideration is the research about self-explanation; an efficient learning strategy where students are prompted to verbalize their thoughts and explanations about the target domain to make knowledge personally meaningful. Previous research [12] found that the amount of self-explanation that students generated in a computer environment was suppressed by having learners type rather than speaking. Moreover, some students are natural self-explainers while others can be trained to self-explain [13]. Even when self-explanation is explicitly elicited, it can be beneficial [14] but requires going beyond asking students to talk aloud by using specific reflection prompts [13].

Self-explanation can be viewed as a tool to address students' own misunderstandings [14] and as a 'window' into students' thinking. While it may be early days for accurate speech recognition to be able to highlight specific errors and misconceptions, undertaking carefully-designed tasks can help identify systematic errors that students make. For example, [15] explores how naming and misnaming involves logic and rules that often aid or hinder students' mathematical learning and relate to misconceptions. A lack of mathematical terminology can also be noticed and prompts made to students to use appropriate language as they self-explain.

Finally, speech provides an additional cue for drawing inferences on students emotions and attitude towards the learning situation while they are solving tasks. By paying attention to tone and pitch of speech in conjunction with other auditory signs like sighs, gasps etc., we can provide learners with even more individualized help, for instance, in the form of motivational prompts.

3 The Wizard-of-Oz study

3.1 Methodology

The studies reported on this paper are part of a process of a methodology referred to as Iterative Communication Capacity Tapering (ICCT) for designing the intelligent support for helping students in interactive educational applications [16]. In particular, they relate to both task-specific problem solving and affective support. ICCT is an extension of the well-known HCI wizard-of-oz methodology for the development of intelligent systems that recognises the complexity of educational contexts by advising a gradual reduction (tapering) of the communication between a human facilitator and the students followed by replacing the facilitator by a computer-based system. During the first phase, the facilitator gradually moves from a situation in which the interaction with the student is close, fast, and natural (i.e. face-to-face free interaction) towards a situation in which the interaction is mediated by computer technologies (e.g. voice-over-ip or similar for voice interaction, instant messaging or similar for textual interaction) and regularised by means of a script. On a second phase, the script is crystallized into a series of intelligent components that produce feedback in the same way that the human facilitator did. The gradual reduction of communication capacity and the iterative nature of the process maximise the probability of the computer-based support being as useful as the facilitator's help.In this paper, we are already starting the second phase, i.e. gradually replacing humans by a computer-based system. Experts ("wizards") are not physically near enough to the students to observe them directly, and therefore must observe them by indirect mediated means: the students' voice was heard by using microphones and headsets and their screen was observed by a mirror screen. The wizards did not have direct access to the students' screen (so e.g. could not point to anything on the screen to make a point), could not see the students' face (for facial cues), and could not communicate to students by using body language, only the facilities provided by the wizard-of-oz tools that resemble those of the final system.

3.2 Participants and Procedure

After returning informed consent forms signed by their parents 60 students 9 to 10-year old (Year-5) took part in a series of sessions with the iTalk2Learn platform configured for learning fractions through structured tasks from the intelligent tutoring system Whizz Maths and more open-ended tasks from the exploratory learning environment Fractions Lab. The sessions were designed to first familiarise all students with the environment and then allow them to undertake as many tasks as possible (in a study that has goals outside the scope of this paper). In parallel, we were running the WOZ study by asking two students in each session to work on different computers as described below. In total 12 students took part in the WOZ study but due to data errors we were able to analyse the interaction of only 10 students.

At the end of the session the students who participated in the WOZ study answered a questionnaire and took part in a focus group. Although methodologically students' opinion elicited through questionnaires can be problematic, they can still provide useful metrics that help us gauge students' perception of the intelligent support and influence our decisions in relation to the the overall approach we are taking [17]. We employed a 5-point Likert visual analogue scale with pictorial representations of smileys that children can relate to (see [17,18]) in order to respond to pertinent questions. We were particularly interested in *helpfulness*, *repetitiveness* and *comprehension*, metrics which we have previously identified useful [17]. In this research we added an extra construct that asked students whether they enjoyed thinking aloud.

3.3 Classroom setup

The ecological validity of the study was achieved by following the setup depicted on Figure 1 and Figure 2. The classroom where the studies took place is the normal computer lab of the school in which more computers are on tables facing the walls on a Π-shape and a few are on a central table. This is the place where the WOZ study took place, while, for ecological validity, the rest of the class was working on the other computers. The students were told that the computers in the central isle were designed to test the next version of the system and were thus also responding to (rather than just recording as the rest of the computers) their speech. The central isle had two rows of computers, facing opposite directions, and isolated by a small separator for plugs etc. In the central isle the students worked on a console consisting on a keyboard, a mouse, and a screen. Usually, those components are connected to the computer behind the screen; for these studies, they were connected to a laptop on the wizards' side of the table. This allowed the wizard to observe what the students were doing. As the iTalk2Learn system is a web-based system, and all the students see is a web browser, the operating system and general look-and-feel of the experience was equivalent to the one that the rest of the students were using. When the wizards wanted to intervene, they used the iTalk2Learn WOZ tools to send messages to the student's machine. These messages were both shown on screen and read aloud by the system to students, who could hear them on their headset.

3.4 The wizard's tools

In line with the ICCT methodology mentioned above, the wizards restricted their 'freedom' in addressing the students by employing a pre-determined script in which the expected interventions had been written. Figure 3 shows a high-level view of this script, the end-points of which require further decisions also agreed in advance in a protocol but not shown here for simplicity. In this study, we limited ourselves to written interventions that could be selected from an

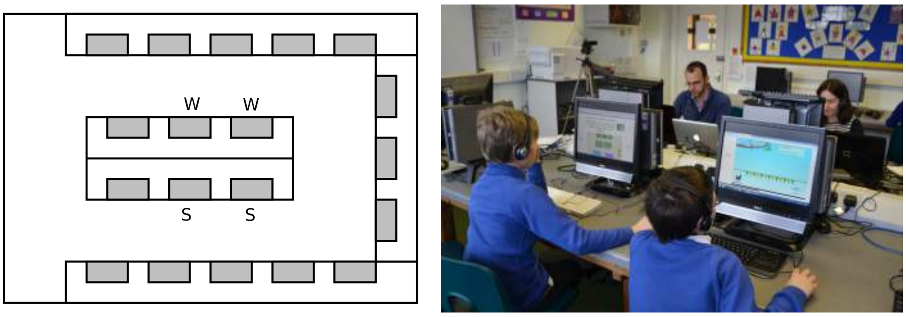

Fig. 1. The classroom. Most students are facing the walls on a Π-shape. The Wizard-of-Oz studies took place on the central isle while the rest of the students are working on a version of the system that only sequences tasks and provides minimal support.

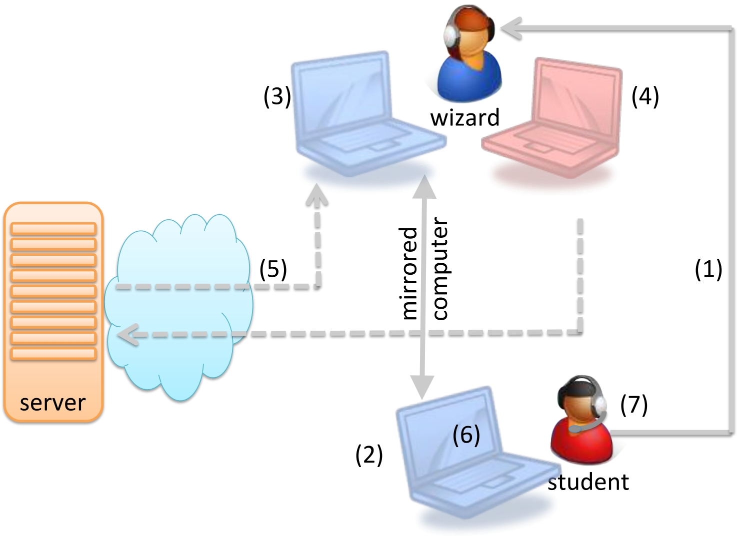

Fig. 2. Wizard-of-oz setup. Each student speaks on a headset (mic) that is connected to the wizard's headphones (1). The student interacts with a console (i.e. keyboard, mouse, screen) that is connected to a laptop on the wizard's side (2,3) so as the latter can witness their interaction. The wizard can send messages (4) by using specially designed wizard tools. These messages arrive to a server and subsequently to the mirrored laptop) (5) where they can be seen (6) and heard (7) by the student.

online document appropriate for being read aloud by the system. There were no other kinds of interventions (such as sounds, graphical symbols on screen etc.). The intervention had a set of associated conditions that would fire them thus resembling very closely the system under development.

Feedback type	Example
AFFECT	It may be hard, but keep trying. If you find this easy, check you work and change the task
TALK ALOUD	Remember to talk aloud, what are you thinking? What is the task asking you?'
TALK MATHS	Can you explain that again using the terms denominator, numerator?
PROBLEM SOLVING	You can't add fractions with different denominators.
REFLECTION	What did you learn from this task? What do you notice about the two fractions?

Table 1. Examples of feedback types

4 Results

In total 170 messages were sent to 10 students. The raw video data was analysed by a researcher who categorised the feedback messages. The researcher noted whether the feedback was directly related to what the student had said; and additionally whether the student reacted immediately, after a delay, or not at all. Another researcher went through the categories and any discrepancies were discussed and resolved before any analysis took place.

Table 1 shows the different types of messages sent to students. It can be seen that most frequent messages were reminders to talk aloud (68). This was followed by problem-solving feedback (55), and feedback according to students emotions (31). The least frequent messages relates to reflection (11) and using maths terminology (5).

By design, all of the reminders to use mathematical terminology were based on students' speech. In 16 of 31 cases, the feedback immediately related to what the student had said concerned emotions. It is not surprising that student reminders to talk aloud were not sent according to what the student had said, as they were provided when the student did not speak.

Students' reactions to the different feedback types can be seen in Table 2. Students reacted to all of the reminders to use mathematical terminology (100%) by repeating their sentence and making an effort to be more precise and use mathematical language such as 'fraction' (instead of 'that') and numerator, denominator (instead of 'this', 'top number', 'bottom number') as they often do. Additionally, students reacted to all feedback to reflect on the task (100%).This was followed by problem solving feedback (87%), and reminders to talk aloud (82%). The lowest number of reactions occurred after an affect boost (74%).

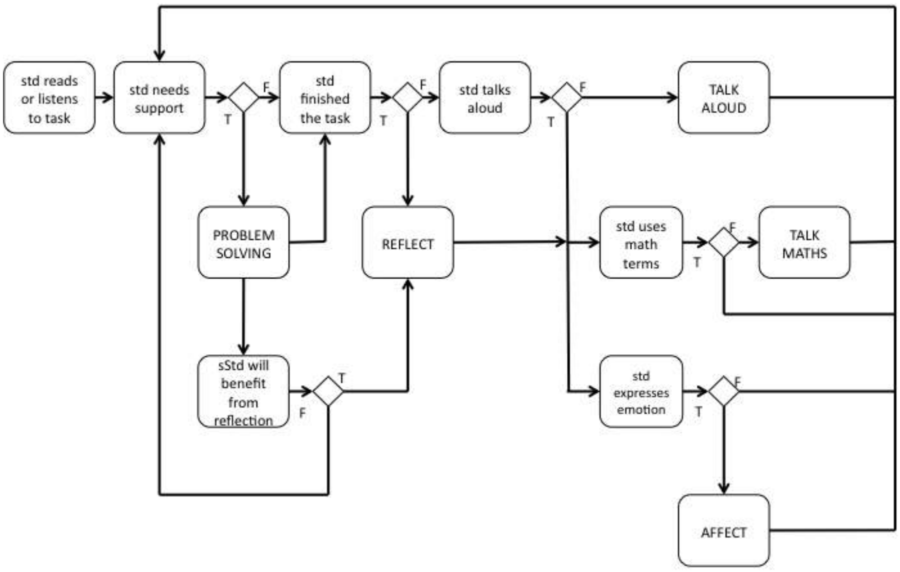

Fig. 3. Flowchart representing the wizard's script for support.

	Was the feedback immediately related to what the student said?		
Feedback type	NO	YES	Total
AFFECT	15	16	31
TALK ALOUD	62	6	68
TALK MATHS	0	5	5
PROBLEM SOLVING	40	15	55
REFLECTION	9	2	11
Total	126	44	170

Table 2. Feedback types, including whether message was sent according to what student said

	Student reacted		Response		
Feedback type	NO	YES	Immediate	Delayed	Total
AFFECT	8	23	19	4	31
TALK ALOUD	12	56	54	2	68
TALK MATHS	0	5	5	0	5
PROBLEM SOLVING	7	48	46	2	55
REFLECTION	0	11	10	1	11
Total	27	143	134	9	170

Table 3. Feedback types and whether student reacted immediate or not

We investigated whether there is any correlation between feedback immediacy and response. There was no significant correlation between provision of feedback immediately after what the student had said, and students' reactions ($r=.18$, $p>.05$). However, the correlation indicates a positive trend on provision of feedback which is immediately following student verbalisation, followed by a reaction. Especially when we take the problem-solving support into account that does not rely on speech but on students' actions with the exploratory environment, there is a more clear correlation ($r=.16$, $p<.05$), as expected because of the several interventions on problem solving that do not necessarily require an immediate reaction from the student, but either to observe or think something that may not be directly observable.

We ran a one-way ANOVA on feedback types with respect to response types by categorising them as follows: affect-related, talk-aloud prompts, and learning-related (includes prompts to elaborate on the terminology, as well as specific problem solving feedback and reflection prompts). There was a significant effect of feedback type on immediacy of the response in that students were less likely to respond immediately after affect prompts $F(2,167)=4.05$, $p<.05$. We discuss this in light of the focus group and student questionnaire results in Section 5.

4.1 Student Questionnaire

Consistent with the findings in the literature and our previous studies, the students responded positively on the questionnaire and did not have any difficulties answering using the smiley visual analogue. Figure 4 depicts boxplots of the answers to the questionnaires that overall are quite satisfactory. One of the reasons for providing the questionnaire is that despite its limitations it can help us identify any negative perception that students may have, that tends not to be voiced in one-to-one or focus groups. As such, it was encouraging to see that all students find the feedback 'somewhat' or 'very' understandable. They also did not seem to find it repetitive, which is positive given the effort that we put in the design and the alternative texts that wizards have in their disposal even for the same meaning. We will discuss in more detail in the next Section why some students may not have found the support so helpful (even if we know that generally this is the lowest reported metric because what people, and young children in particular, perceive as helpful might not necessarily be aligned with what the pedagogical design suggests). The next section also discusses that the variety of answers with respect to whether students liked thinking aloud, even if positive overall, relates to individual differences and preferences.

5 Discussion

Designing intelligent educational systems is a complex endeavour that requires a holistic approach to both the system's behaviour and crucially the interaction with the student. The emphasis here is on interaction since (apart from the difficult task of providing problem-based support at the right time and level) it

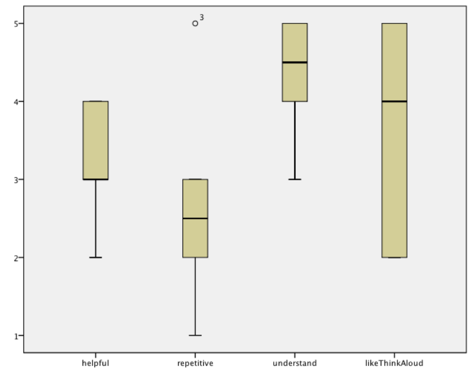

Fig. 4. Box plot of student answers

is important (especially during a WOZ study) to give student the appearance of an interaction that can help them believe both that it is adaptive to their interaction and that there is a benefit in, for example in our case, the effort they are putting in talking-aloud. The issues therefore are intertwined and relate both to the speech recognition capabilities (which here were simulated by wizards and thus had reduced error than an actual system) and feedback capabilities (which in our case was through the written and audio prompts). In what follows we categorise in themes the various issues that have emerged during our analysis including the focus group discussion with the students.

5.1 "Now you are talking" - The effect of ecological validity

From a methodological perspective this study confirmed related literature and our anecdotal evidence that younger students are comfortable talking aloud when undertaking problem solving tasks. It also supported our conjecture that an ecologically valid setting would make a difference compared to earlier work in school with small individual or pair of students. In our previous visits students worked in a small room and at a single computer that was set up to receive WOZ messages. Anecdotal evidence showed that in this more artificial setup, students were less forthcoming thinking aloud despite prompts from the computer and/or the researcher. Prompts of the researcher to "Remember to think aloud" or to "Talk to the computer" were met with responses such as "I don't know what to say" or "I don't know how to talk to the computer". Some of the same students but even other ones who, according to the teacher's admission, are generally shy to talk, when in the authentic classroom setting talked more freely. This is in line with the self-explanation literature discussed in Section 2 that says that students are more likely to talk if they believe that this will help them solve the problem, rather than talking to an audience. There is a lot to attribute to the improved interaction with the computer (as compared to the previous, versions several things had improve to get the students a more realistic interaction).

However, while a proper statistical comparison is not appropriate in this case, we are convinced that the main difference was the ecologically valid setting i.e. that the students who were 'wizarded' felt less observed and therefore were more inclined to think-aloud.

5.2 "Sarcasm doesn't become you" - Speech production matters

Students were positive about their experience at the computer overall. They liked receiving feedback as a direct result of their engagement with the tasks, particularly when it was reinforcing desired behaviour.

The students suggested improvements that could be made to the system. In earlier trials students in another school had heard feedback read out by a male voice. They did not like this voice, reporting it to be very brusque suggesting it sounded like they were being reprimanded each time. In order to improve the experience in this trial, an English female voice was used. The students felt that this voice had a sarcastic tone, particularly for feedback such as "well done". Furthermore, the students felt not all messages needed to be read out (particularly the feedback related to trying other problem-solving approaches), or that some could be stated but not shown as a pop-up because these tended to interfere with the flow of their work. This concurs with [13] findings when designing a multimedia environment that supports self-explanation by avoiding the duplication of messages across two different modalities (e.g. text and narration) that uses the same information processing channel. There was also a suggestion that students could choose when they wanted to read or hear feedback by having a button that they could click on when they needed help or wanted to hear what the computer was suggesting to them.

5.3 "Can the headsets look in my head?" - Cognitive load and AI

Most students, during the focus group, stated that they were at ease thinking aloud some even saying that it helped them concentrate. They were amused about the capabilities of artificial intelligence (AI) and provided (sometimes utopian) suggestions about what it can do. A couple however answered negatively in the questionnaire and in the comments raised a concern that they sometimes struggled to explain what they are doing. Although the latter is not a good justification for not thinking aloud (i.e. struggling to explain, hard as it may be, is definitely useful in clarifying one's thinking and contributes to learning as discussed in Section 2) we observed in some occasions that some of the students would get cognitively overloaded and lost in their own thinking aloud process in an effort to respond on the system's prompts. As the wizards were avoiding performing deep natural language processing they could not help the students. We observed similar situations in the recordings of the rest of the classroom that was encouraged to think aloud (in order to record their voices for training purposes of the ASR system) unaided. In a classroom of course this meant that a teacher or teacher assistant was able to provide more support. The focus group discussion revealed that a combination of factors might be affecting the students' perception on what they can or cannot do. Although all students undertook the same briefing, the fact that a computer can help them when talking aloud is not necessarily something that sinks in quite easily for all of them. This is exemplified by the student who muttered the title of this subsection to himself surprised by the feedback he received when after reflecting on a task he was prompted to repeat using mathematical terminology.

5.4 "Hmm—let me rephrase that" - Pronouns 0 — Maths 1

We observed that most of the times the feedback served its function in that it modified the students' behaviour. The 143 out of 170 feedback prompts where student reacted is indicative of that. Albeit few, we are particularly interested in the prompts that requested rephrasing using mathematical terminology. We do not have the data to support further the conjecture but it seems to be a powerful enough prompt to help the students think more carefully what they are saying. One of the reasons for the low number of this type of feedback messages is that it was only needed once per student. The prompt seems to have been internalised by the students who (at least in one occasion) self-corrected herself replacing demonstrative pronouns with descriptive mathematical entities.

5.5 "Talking the talk" - Individual differences matter

Individual differences matter in learning and interaction. Students' personality, affective characteristics and other factors play a role in their natural propensity to think aloud and/or to talk to an inanimate computer. Among the 12 students two were selected by their teacher on purpose as generally more silent in class and in the lower attainment group. Those received significantly lower feedback prompts related to their speech. We do not have enough data to support this statement but their teacher commented that even the few statements that the system (in the form of the wizard) elicited from them is an achievement. Furthermore, as with other students, we did observe them mumbling to themselves. So even if they were not talking aloud they seemingly engaged in inner-speech, which as mentioned in Section 2, has the potential to help reflection.

6 Conclusion

We presented an ecologically valid Wizard-of-Oz study designed to explore the potential of Automatic Speech Recognition (ASR) to support young students' exploration and reflection as they are working with interactive learning environments. The promising results indicate that, compared to our previous one-to-one settings, there is potential in expecting young students to think-aloud while interacting with educational technology especially if they see value by receiving support. Furthermore, even rudimentary ASR and decision-making mechanisms, as the ones presented here and simulated by our wizards, have the potential to support at the very basic level reflection on the learning task and on the use of the domain terminology. Additional information on affective state derived by cues either in the transcript or in the audio stream can also help in adapting problem solving support [19]. We conducted a cursory analysis on top-level categories that indicated a positive trend on students' more likely reaction on those feedback messages that are immediately following students verbalisation, evidence of a natural interaction. While students did not seem to react (at least in an observable manner) on affect-related prompts, the relationship between affect and feedback provision is very difficult to tease apart [20] and requires further research that was out of the scope of this paper.

We observed the positive reaction of the students to the (simulated) 'system' and collected their comments that feed to our next iteration. In particular, we identified a strong link between the quality of the speech production and the overall interaction with students' perception of the system and subsequent talk aloud. Our immediate steps in an effort to make the system more fluid is to improve the interaction by separating prompts that can only be read to students in contrast to those that have to be shown as well. Future work on the interaction will focus on how interruptive the messages are and carefully selecting the timings of the reflection prompts. We are leaning towards post-task reflection that is more natural and still has the capacity to contribute to student learning. In that sense we can utilise ASR in a non-intrusive manner and open the window to speech-enhanced intelligent support.

Acknowledgments. This research has received funding from the EU 7th Framework Programme (FP7/2007-2013 grant agreement n 318051). We would like to thank the school, students and teachers who took part in the study.

References

1. Flavell, J.H., Green, F.L., Flavell, E.R., Grossman, J.B.: The Development of Children's Knowledge of Inner Speech. Child Development **68**(1) (1997) 39–47
2. Schoenfeld, A.H.: What's all the fuss about metacognition? In Schoenfeld, A.H., ed.: Cognitive science and mathematics education. LEA (1987) 189–215
3. Kafai, Y.B., Harel, I.: Learning through consulting: When mathematical ideas, knowledge of programming and design, and playful discourse are intertwined. In Harel, I., Papert, S., eds.: Constructionism. Ablex Publishing Corporation (1991)
4. Freudenthal, H.: Major problems of mathematics education. Educational Studies in Mathematics (1981) 133–150
5. Canny, J.: The future of Human-Computer interaction. Queue **4**(6) (2006) 24–32
6. Mostow, J., Aist, G.: Evaluating tutors that listen: An overview of project listen. In Feltovich, K.F..P., ed.: Smart Machines in Education. MIT/AAAI Press (2001) 169–234
7. Askeland, M.: Sound-based strategy training in multiplication. European Journal of Special Needs Education **27**(2) (2012) 201–217
8. Borasi, R., Siegel, M., Fonzi, J., Smith, C.: Using transactional reading strategies to support sense-making and discussion in mathematics classrooms: An exploratory study. Journal for Research in Mathematics Education **29** (1998) 275–305

9. Mercer, N., Sams, C.: Tecahing children how to use language to solve maths problems. Language and Education **20**(6) (2007) 507–528
10. Mayer, R.E., Moreno, R.: Nine ways to reduce cognitive load in multimedia learning. Educational Psychologist **38**(1) (2003) 43–52
11. D‘Mello, S.K., Dowell, N., Graesser, A.: Does it really matter whether student‘s contributions are spoken versus typed in an intelligent tutoring system with natural language? Journal of Experimental Psychology; Applied 2011 **17**(1) (2011) 1–17
12. Hausmann, R.G.M., Chi, M.T.H.: Can a computer interface support self-explaining? Cognitive Technology **7**(1) (2002) 4–14
13. Roy, M., Chi, M.T.H.: The self-explanation principle in multimedia learning. In Mayer, R.E., ed.: Cambridge handbook of multimedia learning. New York: Cambridge University Press (2005) 271–286
14. Chi, M.: Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In Glaser, R., ed.: Advances in instructional psychology. Mahwah, NJ: Lawrence Erbaum Associates (2000) 161–238
15. Furani, H.A.: Misconceiving or misnaming?: Some implications of toddlers' symbolizing for mathematics education. Philosophy of Mathematics Education Journal **17** (2003)
16. Mavrikis, M., Gutierrez-Santos, S.: Not all wizards are from Oz: Iterative design of intelligent learning environments by communication capacity tapering. Computers & Education **54**(3) (April 2010) 641–651
17. Mavrikis, M., Gutierrez-Santos, S., Geraniou, E., Noss, R.: Design requirements, student perception indicators and validation metrics for intelligent exploratory learning environments. Personal and Ubiquitous Computing **17**(8) (May 2013) 1605–1620
18. Read, J., Macfarlane, S.: Endurability, engagement and expectations: Measuring children's fun. In: Interaction Design and Children. Volume 2. (2002) 1–23
19. Janning, R., Schatten, C., Schmidt-Thieme, L.: Feature analysis for affect recognition supporting task sequencing in adaptive intelligent tutoring systems. In: EC-TEL 2014. (to appear)
20. Porayska-Pomsta, K., Mavrikis, M., Pain, H.: Diagnosing and acting on student affect: the tutor's perspective. User Modeling and User-Adapted Interaction **18**(1) (February 2008) 125–173