

# Interaction Analysis in Online Maths Human Tutoring: the case of Third Space Learning

Anonymised authors  
Anonymised institutions

**Abstract.** This ‘industry’ paper reports on the combined effort of researchers and industrial designers and developers to ground the automatic quality assurance of online maths human-to-human tutoring on best practices. We focus on the first step towards this goal. Our aim is to understand the largely under-researched field of online tutoring, to identify success factors in this context and to model best practice in online teaching. We report our research into best practice in online maths teaching and describe and discuss our design and evaluation iterations towards annotation software that can mark up human-to-human online teaching interactions with successful teaching interaction signifiers.

**Keywords:** online tutoring, annotation software, best practices

## 1 Introduction

Improving the quality and quantity of the teacher workforce is one of the significant aims of educational research including the field of AIED. Recently, a report from the Unesco Institute for Statistics (2016) stated that almost 69 million teachers need to be recruited globally by 2030 to ensure that international pledges on primary and secondary education are to be kept. The authors of the report argue that there are about 263 million children without a primary and secondary school to attend and there need to be major changes in teacher recruitment to overcome these massive shortages. These shortages are more significant in the subject areas of Science, Technology, Engineering, and Mathematics (STEM) and there is a need to improve the quality of existing teaching in all subject areas for the overwhelming majority of the globe. The concern with teacher quality is partly driven by a growing recognition, fueled by accumulating research evidence, of how critical teachers are to student learning (Hattie, 2012; Sadler, Sonnert, Coyle, Cook-Smith, & Miller, 2013). AIED technologies have great potential to tackle these challenges.

One potential solution to the lack of high-quality teaching at scale is online human teaching approaches. Online human teaching approaches can leverage digital technology to create a successful pedagogical model so that every child, irrespective of wealth or location, can access high-quality teaching. Third space learning (TSL) is a leading online teaching company focusing on primary Math education and the provision of an audio and shared working space platform for teacher to learner interactions.

TSL’s team of teachers, technologists and entrepreneurs based across London, India and Sri Lanka collaborate to design and deliver human one-to-one teaching experiences to children in schools across England. A key challenge in this endeavour is the need to ensure that all online teaching sessions are high-quality. This quality assurance process requires many human evaluators who listen and watch a sample of each tutor’s interactions every week. If there are any causes for concern, the teacher is evaluated more frequently and offered training.

The overarching goal of the research presented here is to automate this evaluation process through the careful design and implementation of AI. We report our early work as we progress towards this goal, in particular we present work towards annotation software to mark up TSL online teaching interactions with successful teaching interaction signifiers. Our aim is to understand the largely under-researched

challenges and success factors in this context and to model best practice in online teaching. In this paper we report our research into best practice in online Math teaching and we describe and discuss our early design and evaluation iterations towards the annotation tool described above.

## **2 Best practice of online Maths tutoring**

Best practice may be framed in different ways. Best practice measurements very often serve as proxies for effectiveness, hence they appear to be generally measured at the institutional level rather than the individual learner level. “Institution” in this case refers to a group of students, teachers and administrators who are together for a common purpose hence; it could be a school, a department of a school, or a programme in a department, for example.

At the individual tutor level, best practice of online tutoring is a combination of various verbal and non-verbal interactions that would lead to student learning. In this sense, best practice is an inseparable part of student learning. Student learning is also a very broad term, which encompasses academic achievement; engagement in educationally purposeful activities, which lead to the development of certain intellectual attributes, such as the acquisition of desired content knowledge, and the development of skills, competencies and self-efficacy.

The list of dependent factors as proxies of student learning is extensive. However, three main themes emerge and must be taken into account in defining best practice:

1. The cognitive domain involves knowledge, understanding and skills about the studied content.
2. The metacognitive domain encompasses the acquisition of knowledge and skills related to one’s own learning, in other words the learners’ knowledge and understanding of their own learning.
3. The affective domain involves learners’ capacities to deal with their emotions, such as attitudes, locus of control, self-efficacy and interest, for example.

### **The Cognitive Domain**

The most important predictors of learner success are often related to the cognitive domain (Stankov, 2009). Cognitive attainment can be measured in many ways, but in most cases measurements are of what learners can achieve in terms of success at solving a maths problem, writing an essay, completing a multiple choice assessment or whatever form of assessment is in place. Few measurements of cognitive attainment consider the process through which learners achieve success, rather they evaluate success in this somewhat decomposed manner. AIED technology can and has considered the measurement of learners’ progress towards attainment. For instance, The Andes Intelligent Tutoring System has rules that are based on a cognitive model of knowledge acquisition that has been developed by analysing protocols of physics students completing example problems. Similarly, Cognitive Tutors maintain a cognitive model instantiated as a system of if-then production rules that can generate the multiple possible correct or incorrect solution steps that a student might take.

### **The Metacognitive Domain**

Metacognition can be broadly referred to as any knowledge or cognitive process that refers to, monitors, or controls any aspect of cognition. When confronted with an effortful cognitive task it is those with greater Metacognitive abilities who tend to be more successful. The Metacognitive domain encompasses actions such as effective goal setting, self-explanation of one’s goals and cognitive processes, reflecting on one’s own learning activities and regulating them. It is argued that those students who have mastery goals (Locke & Latham, 2006); skills clearly and concisely express their goals and thinking process (Chi & Wylie, 2014), a growth mindset (Dweck, 2007) a

reflective practice (King & Kitchener, 1994) often outperform their peers.

### The Emotional Domain

There is a rich literature on the relationships between emotions and learning (see review in Porayska-Pomsta et al., 2008). One model previously used in the design of AIED systems is the cognitive-motivational model (Pekrun, Goetz, Titz, & Perry, 2002). This model describes how a student's learning is mediated by their motivation to learn, their learning strategies, and their cognitive resources. Emotions influence each of these constructs in different ways. For example, emotions direct attention and cognitive resources towards an object or a task, and they can trigger, sustains or reduce academic motivation. An emotional experience such as enjoyment of learning can, therefore, direct a student's full attention to the learning task, enhance a student's academic motivation, and enable the student to adopt flexible learning strategies such as elaboration or critical evaluation.

### Annotating Best Practice in Online Teaching

The theoretical background from the three themes emerging from the literature provided an initial grounding for the first design iteration of the TSL annotation tool (illustrated in Figure 1). Teacher and student interactions were coded in the three dimensions of cognitive, metacognitive and affective domains.

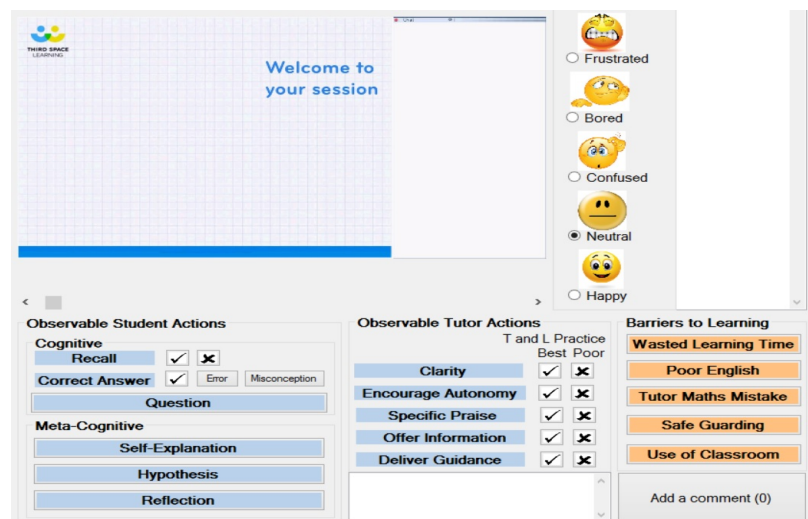


Figure 1. Initial TSL Annotation Tool

The top left corner in Figure 1 presents the tutor's interface to the student. The top right corner has the emotional tags and the list of previously tagged actions. The bottom left corner identifies the observable student actions from the cognitive and the meta-cognitive domains, the bottom center screen has the observable tutor actions, and the bottom right corner has the general barriers to learning items. The tool also has the feature of providing qualitative feedback to the tutor from the evaluator through the addition of a comment at the right bottom corner of the screen

### Prototype evaluation

The overarching aim of this study was to identify observable features of online teaching in order to be able to use them to build a dynamic best practice model for the TSL AI system. The initial annotation tool was evaluated across 20 teaching sessions. These sessions were evaluated by 10 evaluators using the tagging tool. In this section, we analyse their feedback on the tool's potential to capture best practice online teaching according to two research questions:

**RQ1)** What are the teachers' challenges tagging online tuition practice using the tool?

**RQ2)** How can these challenges be addressed to improve the tagging tool's accuracy in capturing observable features of online tuition practice?

### **Participants**

The participants were ten teachers aged between 24 to 41 years. Three of the teachers were recruited from the TSL's tutor evaluator's team, two of them were working as educational researchers and the remaining five were trained teachers enrolled in a masters course in Educational Technology at <anonymised>.

### **Session Tagging Activity**

Teachers were set the task of tagging two sessions each lasting approximately one hour. Before the teachers started the tagging they were all introduced to the tagging tool, trained on how it works and its design decisions (what items should be in the tagging tool and why). Then, they were allowed to interact with the tool and ask any questions that arose as a result. This training session lasted around three hours and teachers were subsequently allocated to tag two of the ten online teaching sessions. Therefore, each session was tagged by two different teachers to be able to compare the reliability of the tagging for each session. At the end of the activity, each teacher was given a form with four questions.

1. Which elements of the session tagger work well, what would make them even better?
2. Which elements were challenging to use, what would you change?
3. What relevant aspects of online teaching were not captured by using the tool?
4. How could we capture these?

Answers provided to these questions were accumulated in text and were qualitatively analysed using thematic coding.

### **Results**

In this section, we present the analysis and results for research question 1: What are the challenges of teachers tagging online tuition practice using the TSL tool?

Two main themes emerged to the first open-ended question that relates to the positive features of the initial version of the session tagger.

- **The tagger was considered as a fast evaluation tool:** Eight teachers mentioned in their responses that the tagger had the potential to be a fast evaluation tool. It was a common feedback that the evaluation of a session might take up to three hours using traditional performance evaluation through open-ended questions. The tagging tool was therefore perceived as an efficient tool for evaluating teachers' performance.
- **The approach of real-time tagging was considered as potentially more accurate than post session evaluation of tuition quality:** Most teachers, six out of ten, thought that the tagging of real-time interactions could lead to more accurate evaluations of tutor performance compared to post-session evaluations of performances. Overall, post session evaluations were considered as more vulnerable to tutor evaluators' bias, compared to the tagging tool.

Regarding the challenging features of the tagging tool five main themes emerged:

- **Too much to focus at the same time. Cognitive, metacognitive, and affective domains as well as barrier to learning aspects are hard to focus on simultaneously:** All of the participant teachers argued that it is not possible to focus on the items from all three domains at any given time. As a result, they appeared to be focussing on certain aspects for a while and then changing to another aspect to tag after a while in a random manner.
- **It is hard to focus on student and tutor actions at the same time:** Again, all of the participants thought that it was confusing to focus on student actions and the

tutor actions simultaneously. It appeared that the teachers were switching from tagging the tutor actions to student actions and vice versa in a random manner.

- **Hard to comprehend and identify some of the tags:** Seven out of ten teachers mentioned that they struggled to understand what exactly some of the tags meant (for instance specific praise, recall, misunderstanding) and what are the examples of these tags. Participants had all joined a training workshop about the annotation tool. Nevertheless, while completing the tagging, they needed some support to identify some of the tags of student and tutor actions.
- **The detection of the affective domain is challenging:** Half of the teachers argued that as the interface does not allow them to see the students while they are working, it was quite challenging to isolate certain nuances between the affective domain stages only from students voices. Although the happy and frustrated states can be identified relatively easily, the confused and bored stages were found challenging to identify.
- **Some user interface and technical challenges:** In addition to the points above some teachers commented on technical or user interface related challenges.

## The Second Design Iteration of the TSL Annotation tool

The results of this first iteration of the TSL annotation tool was the following changes:

- 1) Separate the ‘barriers to learning’ section, which involves evaluation of some general teacher qualities, such as clarity of language. These features are less dynamic and can be judged post-session.
- 2) Student and tutor actions should be coded separately in order to avoid confusion and increase the reliability of the tagging. This might be done through the selection of two equal length sections, one tagged from the student’s perspective and the second from the tutor’s perspective.
- 3) Separate the evaluation of the affective domain from the cognitive and metacognitive domains, because of the intertwined nature of cognitive and metacognitive actions that makes them hard to isolate from each other.
- 4) Provide evaluators with clear definitions and examples of all the tagging items.

Figures 2 and 3 present the revised versions of the annotation tool, with two different interfaces for tagging tutor actions and student actions separately. The tutor actions are categorised in three sections: Question, feedback, guidance and the student actions are categorised in four sections: Question, answer, reflection and self-explanation. These categories are based on the predominant pattern found in the online tuition dialogues of question-answer-feedback cycles where the tutor states a question, the student answers and the tutor follows up with appropriate feedback (Dillon, 1990; Graesser, Person, & Magliano, 1995). Graesser et al. (1995) elaborates on the one-to-one tutorial interactions by providing a five-step dialogue frame where (1) the tutor is asking a question; (2) the student is answering and explaining; (3) the tutor provides feedback on the answer; (4) the tutor and the student are improving the quality of the answer by reflecting on it; (5) the tutor assesses student’s understanding of the answer.

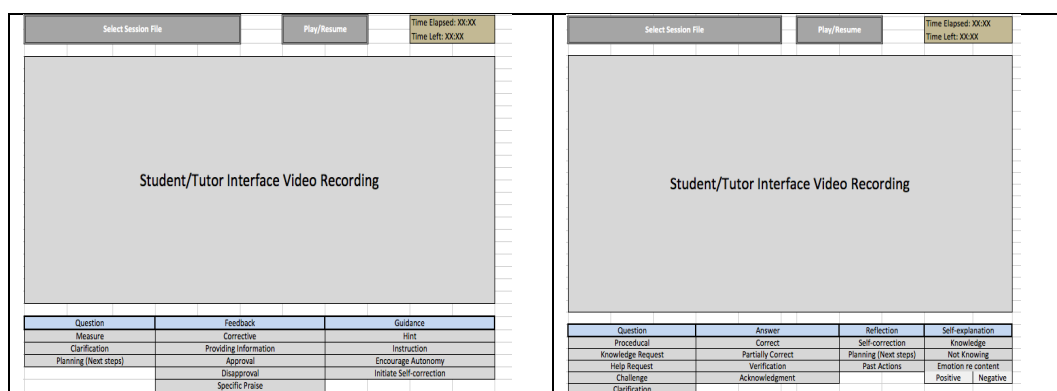


Figure 2: Tutor actions tagging tool for primary Math teaching in online settings

Figure 3: Student actions tagging tool for primary Math teaching in online settings

Similar interactions were identified in more recent research studies (Porayska-Pomsta, Mavrikis, & Pain, 2008). We are now at the stage of evaluating these new tools.

## Conclusions

It is essential that industrial and academic partners work together at an early stage in the design of AIED tools if the research they conduct is to optimally influence the design of the scaled technology. However, this early engagement also presents a dilemma, because, in the absence of an existing validated model of best practice on which to base the AI components, one has to build the model through design iterations with human participants playing the role eventually intended for automation by AI. Some of the problems faced by the human participants may not be an issue for an AI system. For example, the need to separate coding of teacher and learner interfaces. Other problems may be equally or more pertinent for AI systems. For example, the need for a precise definition of the meaning of each tag and the signifier it describes. The heart of the issue is that without this simplification for human evaluators to enable us to evaluate the tool prior to its automation we cannot build and validate our best practice model. A useful by-product of the human level evaluation process may be in its ability to inform the design of interface components for reporting back to teachers the results of the automated evaluation of their teaching sessions. For example, if the combination of teacher and student foci in the interface is confusing for evaluators when they are evaluating a teaching session, then it may well be difficult for teachers to understand when they are trying to use the interface to interpret the quality of their own teaching.

## References

- Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, 49(4), 219-243. doi:10.1080/00461520.2014.965823
- Dillon, J. (1990). The practice of questioning. International series on communication skills, Owen Hargle: London: Routledge.
- Dweck, C. S. (2007). The perils and promises of praise. *ASCD*, 65(2), 34-39.
- Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology*, 9(6), 495-522.
- Hattie, J. (2012). *Visible learning for teachers: Maximizing impact on learning*: Routledge.
- King, P. M., & Kitchener, K. S. (1994). *Developing Reflective Judgment: Understanding and Promoting Intellectual Growth and Critical Thinking in Adolescents and Adults*. *Jossey-Bass Higher and Adult Education Series and Jossey-Bass Social and Behavioral Science Series*: ERIC.
- Locke, E. A., & Latham, G. P. (2006). New directions in goal-setting theory. *Current directions in psychological science*, 15(5), 265-268.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37(2), 91-105.
- Porayska-Pomsta, K., Mavrikis, M., & Pain, H. (2008). Diagnosing and acting on student affect: the tutor's perspective. *User Modeling and User-Adapted Interaction*, 18(1-2), 125-173.
- Sadler, P. M., Sonnert, G., Coyle, H. P., Cook-Smith, N., & Miller, J. L. (2013). The influence of teachers' knowledge on student learning in middle school physical science classrooms. *American Educational Research Journal*, 50(5), 1020-1049.
- Stankov, L. (2009). Conservatism and cognitive ability. *Intelligence*, 37(3), 294-304.