

Adolescents' Self-regulation during Job Interviews through an AI Coaching Environment

Kaśka Porayska-Pomsta¹ and Evi Chryssafidou²

¹ UCL Knowledge Lab, University College London, K.Porayska-Pomsta@ucl.ac.uk

² Metanoia Institute, Faculty of Research Strategy and Innovation, London

Abstract. The use of Artificial Intelligence in supporting social skills development is an emerging area of interest in education. This paper presents work which evaluated the impact of a situated experience coupled with open learner modelling on 16-18 years old learners' verbal and non-verbal behaviours during job interviews with AI recruiters. The results revealed significantly positive trends on certain aspects of learners' verbal and non-verbal performance and on their self-efficacy.

1 Introduction

Despite the importance of social interaction to human quotidian functioning, social skills require substantial training, socio-cultural conditioning and highly developed metacognitive competencies, involving ongoing, targeted self-monitoring and regulation. Emotional self-monitoring and regulation are primary in motivating people to communicate, with emotions also playing a dominant role in learning [1].

Social interaction and emotional self-regulation skills cannot be supported merely through showing or telling people how to feel or behave. Instead, they require access to (i) repeatable embodied experiences in contexts that credibly approximate real-life scenarios, and (ii) opportunities for situated recall and guided scrutiny of the behaviours enacted first-hand by the learners. Delivering the desired learning experiences is challenged by the time-consuming nature of both the set-ups and the support required in this domain.

Currently, two approaches dominate: (i) *vicarious learning*, popular in special needs interventions, e.g. autism, where learners observe recorded or written social stories, which they then discuss with practitioners; (ii) *role-playing* in mock scenarios based on some well-defined rules, often followed by debriefing with practitioners. The vicarious approach permits detailed analysis of the scenarios studied, but not first-hand experiences. Role-playing offers first-hand experiences, but a detailed analysis may be limited by the quality of the data collected. Socially plausible interactions that are supported by AI agents and open learner modelling (OLM) provide a useful alternative to the methods available [2]). They also enable a systematic study of social interactions and of learning support needed.

This paper presents a study which evaluated the impact of an intelligent coaching environment, called TARDIS, involving embodied conversational agents and open learner modelling, on 16-18 years old learners' job interview skills and on their self-efficacy.

2 TARDIS coaching environment

TARDIS is a coaching environment for learners who are at risk of social exclusion through unemployment. It supports learning and exploration of social interaction and self-presentation skills in job interview contexts [3]. It comprises two overarching elements: (i) a *job interview simulator*, for situated rehearsals of learners' job interview skills with AI recruiters (AIRs), (Fig. 1, left) and (ii) an OLM, used to scaffold reflection about learners' behaviours during job interviews (Fig. 1, right).

The simulator comprises an interaction scenarios model, and models of socio-affective and behavioural competencies of the recruiters [3]. All of the models are utilised in an orchestrated data-driven way by the FAtiMA [4] emotion-enhanced planning architecture, which allows to create a variety of emotionally nuanced AIRs [5]. The OLM is based on the pre-existing platform called NOVA [3], which was extended and tailored for use in TARDIS, based on knowledge elicitation with practitioners and annotated interactions of human-human mock job interviews [6]. Learners engage with AIRs verbally, through head mounted ear-/microphone, and through gestures detected by Microsoft Kinect.

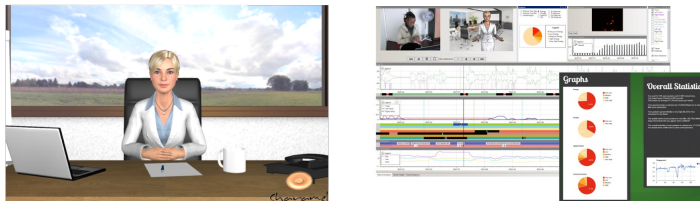


Fig. 1. Left: TARDIS' AIR Gloria; Right: NOVA-supported OLM.

The data recorded through TARDIS tools include videos of learners interacting with the agents, their specific dialogue moves along the interview timeline, as well as learners' verbal and non-verbal behaviours, e.g. head pose and upper-body gestures, and voice quality including pitch, amplitude, energy and duration of their utterances. In the post-interview debriefing phase, these data sources are synchronised with NOVA's analyses and are displayed to the learners and practitioners through the NOVA tool (Fig. 1, right) to facilitate discussion and reflection. The pedagogical set up of TARDIS was designed through a series of studies with learners and practitioners to complement the existing real-world practices used in youth organisations and job centres, and to leverage in a blended way the strengths of both TARDIS and of human practitioners [7].

3 Study Design and Results

TARDIS was evaluated with adolescents aged 16-18 years old (Mean=17.07 years; SD = 0.6), who were identified as at risk of becoming NEETs (*Not in Education, Employment or Training*) after leaving school. The study used a pre- and post-test design, with (i) human-to-human mock interviews, (ii) self-reports of *self-efficacy*, *anxiety* and *quality of performance* as the pre and post baseline measures, and (iii) a control intervention using a web-based programme that is representative of the type of practice currently recommended by job centres in the UK.

Participants. 28 adolescents were divided into intervention (**IG**) and control groups (**CG**) using a randomised matched pairs approach.

Procedure. Mock interview with a human practitioner and a self-report questionnaire were administered to both groups pre- and post- intervention. In both groups the participants viewed their recorded mock interviews and received feedback on three aspects of their performance requiring improvement. In the TARDIS condition, the participants then engaged in three one-hour sessions over three days involving two practice-and-reflection cycles, including (a) learner practicing with AIRs and (b) individualised feedback delivered by a practitioner using the NOVA OLM according to a prescribed procedure. An interaction with an 'understanding' AIR and a 'demanding' AIR provided two test conditions, which were delivered always in the same order. The AIRs behaved in a manner aligned with their respective styles, e.g. involving more or less face-threat. The control intervention involved two web-based training exercises that were completed over two sessions. The first exercise required the participants to choose the correct answers to a series of video recorded interview questions. Unlike in the TARDIS condition, the participants' responses were automatically scored within the programme, followed by suggestions on possible improvements and a second attempt by the participants to improve their scores. The second exercise involved reading through 100 job interview questions, which were accompanied by examples and tips for possible answers.

Measures. Nine measures were used to compare the interventions: (i) *quality of response*: the degree of relevance of learners' responses to interview questions, quality of response structure and the level of content elaboration; (ii) *eye contact*, e.g. rare vs. well-maintained eye contact with the interviewer, (v) *tone of voice*, e.g. monotonous vs. modulated speech, and (vi) *facial expressions*, e.g. indifferent vs. interested. Additionally, learners self-reported on: (vii) *self-efficacy*: their belief that they can do well; (viii) *anxiety*: feelings of apprehension and tension caused by a job interview; and (ix) *quality of performance*.

Results. Two independent annotators coded the human-human interviews using a bespoke schema enhanced with training videos exemplifying the specific behaviours and the recommended scores, and inter-rater agreement was calculated. Respective Kappa's were: $k = 0.64$ for quality of response; $k = 0.3$ for eye contact; $k = 0.3$ for tone of voice; $k = 0.45$ for facial expressions. Note that it is not unusual for post-hoc annotations of voice to yield moderate to low kappa values, indicating that this dimension may be context and culture dependent as well as it

may depend on the individual communicative preferences of the annotator. The examples from the affective computing literature (e.g. [8], [9], [10]) suggest that the results obtained are not out of line with the existing research with respect to voice-based emotion judgements. Similarly eye-gaze is very hard to judge based only on video data and the discrepancies between the two annotator's judgments may be explained by their differing levels of what is acceptable in terms of frequency of saccades, lowering of the gaze or looking up. The fair agreement points to the need for tightening of the guidelines for how to interpret youngsters' eye-gaze patterns in job interview situations and for additional moderation of the annotations – a task which is presently under way.

An analysis of variance (ANOVA) of interviewees' pre- and post- *quality of response* revealed significant improvements for both *intervention (IG)* and *control (CG)* groups, with significantly greater improvements for the IG on two of the most challenging questions (in total there were 16 questions asked): *Q1: 'Why are you applying for this post?'* ($F(1, 26) = 5.45, p < .05$); *Q2: 'Why do you think we should hire you?'* ($F(1, 26) = 6.30, p < .05$). Both groups also improved at post-test on non-verbal behaviours, with the IG showing significantly greater improvements than the CG on *eye contact* ($F(1, 26) = 14.07, p < .01$), *tone of voice* ($F(1, 26) = 13.88, p < .01$), and *facial expression* ($F(1, 26) = 7.5, p < .05$).

ANOVA was also conducted on the three self-reported measures. The results suggest that both groups improved significantly on all of the three measures at post-test (*self-efficacy*: $F(1, 26) = 20.33, p < .0005$; *anxiety*: $F(1, 26) = 13.40, p < .01$; *quality of performance*: $F(1, 26) = 33.33, p < .0001$). However, no intervention effect was found on any of the measures, suggesting that participants in both IG and CG thought that they benefited equally well from their respective experiences.

4 Discussion and Conclusions

Overall, the study results are encouraging, suggesting that although social interaction skills and self-perception improvements can be gained through both types of intervention, the situated practice coupled with OLM-supported feedback from a human may be more effective than the vicarious observation and testing, or tips and advice on impression management. The study highlights the potential role of AIED technologies in supporting social skills acquisition and development of metacognitive competencies in this context, particularly in providing a situated repeatable experience to the learners and opportunities to revisit those experiences post hoc in a manner that is conducive to concrete and systematic reflection and discussion, and ultimately to learning. Future methodological improvements such as a more extensive moderation of the annotations, and an abductive approach will allow to address questions about the impact of the individual aspects of the intervention, e.g. use of TARDIS with and without OLM, on learners' performance.

References

1. Immordino-Yang, M.H., Damasio, A.: We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, Brain, and Education* **1**(1) (2007) 3–10
2. Hoque, M.E., Courgeon, M., Martin, J.C., Mutlu, B., Picard, R.W.: Mach: my automated conversation coach. In: Proceedings of the 2013 ACM international joint conference on Pervasive and ubiquitous computing (UbiComp '13). UbiComp'13, New York, NY, USA, ACM (2013) 697–706
3. Anderson, K., André, E., Baur, T., Bernardini, S., Chollet, M., Chryssafidou, E., Damian, I., Ennis, C., Egges, A., Gebhard, P., Jones, H., Ochs, M., Pelachaud, C., Porayska-Pomsta, K., Rizzo, P., Sabouret, N. In: *The TARDIS Framework: Intelligent Virtual Agents for Social Coaching in Job Interviews*. Springer (2013) 476–491
4. Dias, J., Paiva, A.: Feeling and reasoning: A computational model for emotional characters. In: *Progress in Artificial Intelligence*. (2005) 127–140
5. Chollet, M., Ochs, M., Clavel, C., Pelachaud, C.: Multimodal corpus approach to the design of virtual recruiters. affective computing and intelligent interaction. In: *Proceedings of the Affective Computing and Intelligent Interaction*. hal-01074861, Genve, Switzerland, ACII (2013) 19–24
6. Porayska-Pomsta, K., Rizzo, P., Damian, I., Baur, T., André, E., Sabouret, N., Jones, H., Anderson, K., Chryssafidou, E. In: *Who's Afraid of Job Interviews? Definitely a Question for User Modelling*. Springer (2014) 411–422
7. Porayska-Pomsta, K.: AI as a methodology for supporting educational praxis and teacher metacognition. *International Journal of Artificial Intelligence in Education* **26**(2) (Jun 2016) 679–700
8. Litman, D.J., Forbes-Riley, K.: Predicting student emotions in computer-human tutoring dialogues. In: *In Proceedings Of The 42nd Annual Meeting Of The Association For Computational Linguistics*, East Stroudsburg, PA, Association for Computational Linguistics (2004)
9. Ang, J., Dhillon, R., Krupski, A., Shriberg, E., Stolcke, A.: Prosody-based automatic detection of annoyance and frustration in human-computer dialog. In: *Paper presented at the International Conference on Spoken Language Processing*, Denver, Co (2002)
10. Shafran, I., Riley, M., Mohri, M.: Voice signatures. In: *Paper presented at the Proceedings IEEE Automatic Speech Recognition and Understanding Workshop*, Genve, Switzerland (2003)