

Evaluation of Novice and Expert Interpersonal Interaction Skills with a Virtual Patient

Patrick G. Kenny, Thomas D. Parsons, Jonathan Gratch, and Albert A. Rizzo

Institute for Creative Technologies / USC
13274 Fiji Way Marina Del Rey, CA 90292, USA
{kenny, tparsons, gratch, rizzo}@ict.usc.edu

1 Introduction

Interactive Virtual Standardized Patients (VP) can provide meaningful training for clinicians. These VP's portray interactive embodied conversational characters with realistic representations of a mental or physical problem to be diagnosed or discussed. This research is a continuation of evaluating of our VP "Justina" [2] which suffers from Posttraumatic Stress Disorder (PTSD) from a sexual attack and presents the results of comparing novices, test subjects without medical training, and experts interacting with 'Justina' to find out if they could elicit the proper responses to make a diagnosis and to investigate the topics and questions the novices asked for coverage of the categories and criteria of PTSD as defined in the DSM-IV [1]. It is assumed that novices will perform better than experts, however the main investigation is to gather empirical data and understand why this is true and how this can be used to improve the system. There have not been, to the authors' knowledge, any studies in evaluating experts and non-experts with virtual human characters in the psychological domain.

The subject testing method consisted of recruiting novice participants from USC staff, students and interns. A total of 9 people took part in the study (3 females, 6 males). This was compared against the experts from the previous study, 15 subjects consisting of, medical students, psychiatry residents and fellows (6 females, 9 males). The subjects filled out pre and post questionnaires, the dialog exchange is transcribed and classified into the PTSD categories.

1.1 Data Analytics

For our analyses we focused on cohort (clinicians and novices) differences related to effective diagnostic interview skills. The keys aspects of the interview included differences between clinicians and novices on establishment and maintenance of rapport, attempts to gather information about the VP's problems and attempts at detailed inquiry to gain specific and detailed information from the VP. Question/response composites (VP_QR') were developed to reflect the shared relation existing between the responses of a VP and of DSM IV TR-specific Questions (from both cohorts) that are necessary for differential diagnosis. For the VP_QR' scores, we first calculated eigenvalues via least squares procedures and separate composite measures were created for each observation. The resulting weights were used in conjunction with the original variable values to calculate each observation's score. The VP_QR' scores were standardized according to a z-score. The data analysis was completed in three stages. First, the reference distribution is a correlation of each cluster of questions

(from the clinicians) making up a particular DSM PTSD Category with each (corresponding) cluster of responses from the VP representing the same DSM PTSD Category. The second was the same as the first but for novice questions and VP responses representing the same DSM PTSD Category. In the third stage, the differences between cohorts were assessed to see whether significant differences existed in each cohort's reference distribution of questions response pairs—from the VP representing the same DSM PTSD Category.

1.2 Results and Conclusion

The data for the 9 novices was compared to that of the 15 experts in the previous study. There were less questions asked in the communications (Novice questions: 24.13, VP responses 44.26; Expert: 34.00, VP responses 40.44), for the novices than for the experts. This implies that the novices were asking questions about general things and not specific criteria to help make a diagnosis and the clinicians were better able to maintain rapport than the novices. The novices kept revisiting the same topic and questions more than the experts, and spent a large amount of time on the incident rather than the diagnosis criteria. Additionally findings revealed a significant difference between cohorts on PTSD Categories A(Trauma Event), B(Re-experience event), C(Avoidance), E(Duration), and G(Rapport). No significant differences were found for categories D(Arousal) and F(Life Effects), Table 1.

Table 1. Ratios for 15 Experts and 9 Novices

Category	Mean Expert	Mean Novice	Std. Dev. Expert	Std. Dev. Novice	F-Value
PTSD_A_rto	9.03	0.18*	3.94	0.06	44.50
PTSD_B_rto	7.04	2.15*	3.33	1.42	17.23
PTSD_C_rto	6.21	-0.26*	1.64	0.10	136.45
PTSD_D_rto	1.08	1.66	0.50	2.43	0.82
PTSD_E_rto	-0.13	0.21*	0.08	0.14	57.61
PTSD_F_rto	0.69	1.20	0.42	1.18	2.34
PTSD_G_rto	38.30	-14.89*	10.72	1.80	214.05
PTSD_H_Q	3.67	2.44	3.66	2.07	0.83

*Significance at the .05 level.

The system could be improved by adding additional questions for Category D and since there was not a significant difference in F, the clinicians could use a bit more training here. It was anticipated that the novices would not do as well as the experts; however this was one of the first comparative analysis between novices and experts for VP's in the psychological domain.

References

1. DSM, American Psychiatric Association: Diagnostic and statistical manual of mental disorders, text revision (DSM-IV-TR), 4th edn. American Psychiatric Press, Inc., Washington (2000)
2. Kenny, P., Parsons, T.D., Gratch, J., Leuski, A., Rizzo, A.A.: Evaluation of Justina: A Virtual Patient with PTSD. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 394–408. Springer, Heidelberg (2008)