A Cautionary Note on Personality (Extroversion) Assessments in Child-Robot Interaction Studies

Paul Baxter, Tony Belpaeme
Centre for Robotics and Neural Systems
Plymouth University, U.K.
{paul.baxter, tony.belpaeme}@plymouth.ac.uk

Abstract—The relationship between personality and social human-robot interaction is a topic of increasing interest. There are further some indications from the literature that there is an association between personality dimensions and various aspects of educational behaviour and performance. This brief contribution seeks to explore the single personality dimension of extroversion/introversion: specifically, how children rate themselves with a validated questionnaire in comparison to how teachers rate them using a relative scale. In an exploratory study conducted in a primary school, we find a non-significant association between these two ratings. We suggest that this mismatch is related to the context in which the respective ratings were made. In order to facilitate generalisation of personality-related results across studies, we propose two general reporting recommendations. Based on our results, we suggest that the application of personality assessments in a child-robot interaction context may be more complex than initially envisaged, with some dependence on context.

I. Introduction

The role of personality in human-robot interaction is becoming of greater interest in the field, as attempts are made to increase the adaptability and personalisation of the robots. For example, preference has been found in a rehabilitation context for a robot that matches one's own personality [1]. Similarly, in children, robots that take into account the personality of the interacting child (e.g. if shy) can adapt its behaviour accordingly to promote interaction [2]. In our research, we are generally interested in having robots adapt to children within interactions in order to facilitate some outcome such as learning or behaviour change, e.g. [3]. As a trait upon which adaptation can be based, personality is therefore of interest.

There are a number of problems with administering lengthy questionnaires to children, particularly those related to personality assessments. Primarily, these include the level of concentration required for completion, and the conceptual level of the questions (specifically if abstract, or relating to life experiences that may not be typical for children). For this reason, there have been a wide range of development and validation efforts to produce short-form questionnaires in a range of languages. Part of this validation process frequently involves examining the convergence of personality ratings between parents, teachers and the children themselves, with high agreement being used to support validity, e.g. [4]. One such effort is an abbreviated Junior Eysenck Personality Questionnaire [5], which attempts to characterise four dimensions of extroversion, neuroticism, psychoticism, and (questionnaire)

reliability using 24 questions. This was validated with children in the age range 13–15, although related work showed valid application to children of a slightly younger age [6]. In the present study, we employed a short-form version of a five-factor model questionnaire that has been validated with children [4]: the BFQ-C.

We focus specifically on the single dimension of extroversion. Prior work has, for example, suggested that extroversion is positively associated with verbal-imagery-based learning in children [7], and with help-seeking behaviours (self-regulated learning) in adults [8]. These make it a dimension of interest to our educational context. Extroversion is also suitable as a characteristic of interest since it is a dimension (extroversion vs introversion) that appears in a range of human personality theories (for example both the Eysenck and five-factor 'big 5' models).

In this study, we examine the relationship between self-rated scores of extroversion with teacher ratings of relative extroversion. As a secondary consideration, we also consider the possible relationship to learning outcomes in a subsequent collaborative learning task with a social robot, although this is not the focus of this paper. The work described here accords with our wider goals of ethologically-appropriate and valid empirical investigations for child-robot interaction in educational contexts [9]. First we introduce the exploratory study (section II), before interpreting the results (section II-B) as suggesting that care must be taken in considering the context of the personality assessment (section III).

II. Exploratory Study

As an exploratory study, we do not propose hypotheses. However, from the discussion above, we may venture the predictions that the child self-ratings of extroversion and the teacher-ratings of the same will be positively associated (reflecting that the teachers know the children), and that there will be a positive association between ratings of extroversion and learning outcome. In the following, we assess whether the data provide any support for these predictions.

A. Setup and Method

The study was conducted in two primary schools in the U.K. 38 children, aged 7–8 years old took part (22 boys, 16 girls). The study was run in accordance with a protocol approved by the Plymouth University Faculty of Science and
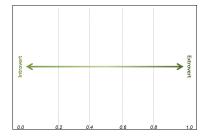
Fig. 1. The teacher scale for relative introversion/extroversion child ratings. Teachers were instructed to write the names of the children on the sheet. Numbers in italics (not displayed to the teachers) indicate coding of the position of the names on the sheet: names on (or next to) the dotted were assigned the score shown; numbers in the space between dotted lines were assigned an intermediate score (e.g. 0.3, 0.5, ...).

Question	Pearson Correlation	
	Self-rating	Teacher-rating
Q1	0.4994	0.1207
Q2	0.1845	-0.2115
Q3	0.4734	*0.3090*
Q4	0.4141	-0.1541
Q5	*0.3112*	0.0076
Q6	0.3872	0.1234
Q7	*0.3187*	0.2414
Q8	0.5706	-0.0226
Q9	0.5675	0.1494
Q10	0.1577	0.1348
Q11	0.4890	0.0900
Q12	0.5463	-0.1016
Q13	0.3937	0.1888

Technology ethics board. An opt-out consent was obtained from all parents/guardians of the children, with separate opt-in consent for image/video recording (not used in the present paper). All children were permitted to withdraw from the study at any point upon request. The experiment took place towards the end of the school year, meaning that the teachers had spent at least the majority of a school academic year with the children.

The visit to each school began in the morning: after initial attendance check, the experimenters were introduced to the class. They informed the children of the purpose of the visit: to play a sorting game with the robot and to fill in some questionnaires (both knowledge pre/post tests and the personality questionnaire). The extroversion scale questionnaire was then administered to the children as a group in the classroom, led by the teacher: independent completion was instructed (and enforced) by the teacher (i.e. prevention of copying).

Separately, the teacher was briefed on their rating of the children's extroversion. On the single dimension of extroverted to introverted, represented on a single sheet of paper (figure 1), the teachers were asked to place the children in their class in relation to one another, based on a similar rating scheme as used in [10]. This scale was therefore an explicitly subjective and relative rather than a subjective and independent measure of this personality characteristic. The intention was to examine the correlation between teacher-ratings and self-ratings rather than a direct comparison of scores.

Through the rest of the day, the children were brought one-by-one into a separate room in the school, where they completed a pre-knowledge test on carbohydrates, engaged in a sorting task with a Nao humanoid robot (Aldebaran Robotics) on the topic of carbohydrates for five minutes, and then completed a post-knowledge test (different pre and post tests, counterbalanced between individuals, not containing images used on the interaction).

B. Results

A qualitative inspection of the data does not suggest any strong relationships between child self-rating of extroversion, the teacher-rating of the same, and learning outcome (figure 2). The correlation between learning outcome, as measured by pre- to post-test score change, and both self-rating (r=0.030, p=.857, n=38) and teacher-rating (r=0.029, p=.863, n=38) is not significant, with very low effect sizes. Due to the incidental nature of learning outcome for the present contribution, we do not consider it further, other than to note this lack of significant association.

Of perhaps more unexpected nature is the low (non-significant) correlation between the teacher-rating and the child self-rating (r=0.142, p=.142, n=38, figure 2(a)). This indicates that there was weak agreement between the children and their teachers, despite spending extended periods of time with each other (i.e. the school days).

Examining the correlations between the data obtained on a single question basis further supports the observation that there is at best only a weak link between the self-ratings and the teacher-ratings. Firstly, as would be expected, there is generally a high number of positive correlations between the individual question responses and the overall self-rating (table I). Secondly, however, this positive relationship is not reflected in the correlation of self-ratings to the overall teacher-ratings. Only for one question (Q3: "I like to move and to do a great deal of activity") is there a moderate positive correlation (though not quite significant, p=0.059, n=38). Interestingly, this positive correlation between physical activity and extroversion has been found in children of this age [11], indicating some (limited) support for the idea that the teachers do have some familiarity with the children, and that there is divergence between the ratings in spite of this.

A further result of interest is related to Q8 ("I like to talk with others"). There is a strong positive correlation between the overall self-rating and the response to this question (r=0.571, p<.001, n=38), but there is no correlation between the response to this question and the teacher-rating of extroversion (r=-0.023, p=.893, n=38). Assuming that willingness of children to speak with others is likely to be one of the more apparent characteristics of children to their teachers, this lack of association is perhaps surprising.
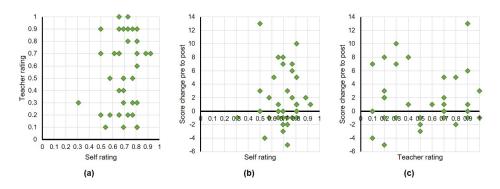
Fig. 2. Raw data scatter plots showing the relationship between the metrics of child self-report extroversion (normalised scale), teacher-reported child extroversion (normalised scale), and learning outcome (*post-test score - pre-test score*): (a) self-rating versus teacher-rating; (b) self-rating versus learning outcome; and (c) teacher-rating versus learning outcome.

III. DISCUSSION

The results of this exploratory study suggest that there is a difference between the way the children see themselves in terms of extroversion and the way their teachers see them. Extroversion in this context seems to be a particularly relevant characteristic to explore in this way given its overt behavioural component. These results are consistent with previous observations that there is little agreement between child self-ratings and teacher-ratings (although parent-ratings fared better) [10], although the present study extends these by significantly extending the number of subjects involved. Indeed, the present results also seem to be in accordance with the results from the questionnaire validation itself, which suggested a non-significant association between teacher ratings and child self-ratings for extroversion, but only for younger children [4].

We, in this and other work, examine interactions between children and robots in school environments. We are therefore interested in characterising various aspects of this context in particular.

One consideration having an effect on these results may be the environment in which the study was conducted, and the relationship between this and the teacher as involved observer. The teacher interacts with the children only during the school day (with the type of educational environment itself providing potential biases of child behaviours and teacher interpretations thereof), and would typically not do so outside of the context of school. The children are naturally not constrained by school alone, and as such will have a broader experience upon which they base their personality self-assessments. It is thus perhaps not surprising that there is an apparent mismatch between the children's perception of themselves (albeit on only one personality sub-scale) and that of their teachers. The question then becomes, which assessment (child or teacher) is more relevant to school performance? This is only speculation, but the results provide a basis for further empirical exploration.

There are a number of issues, both general and specific, with the present study. Firstly, we were limited in our examination of only one sub-scale of personality as characterised by the 5-factor model. Furthermore, while a questionnaire validated with children was used, the three-item scale we employed

(for reasons of clarity for the children) is relatively coarse, thus limiting the resolution of the measure. Nevertheless, the wide range of responses obtained (see Appendix) suggests that inter-personal variability was still discernible. Secondly, we were comparing self-ratings from the questionnaire with relational ratings from a third party (the teacher). Being relational, this explicitly rated the children with respect to one another: our prediction that it is reasonable to examine the association of the two measures is clearly not borne out by the results. While we interpret this as a context effect, there is naturally the possibility that it is our comparative measures approach that is flawed. The mis-matching results nevertheless remain to be explained, and thus still in our view constitute a reason to be wary of the self-rating (or other-rating) measure alone. Thirdly, compared with personality questionnaire validation exercises (with participant numbers typically in the multiple hundreds), our sample size (n=38) is relatively small. Given that at large sample sizes moderate to small correlation coefficients can become statistically significant, it is possible that a more extended study would find that our results were also significant. However, the low effect size (r=0.142) still suggests the lack of a straightforward positive association between self- and teacher-ratings. Finally (and in general), there are also a number of issues related to the administration of questionnaires to children, e.g. [12], as a result of effects such as social desirability, thus calling into question the reliability of such methods. While validation of the questionnaire with the appropriate subject group (i.e. children in this case) can mitigate this effect, it is necessary to remain cautious.

While these issues naturally reduce the potential power of the results obtained, we believe that there are still a number of pertinent points that are raised by this study. Generally, and this is of course an important consideration for any empirical investigation, how can we be sure that we are measuring what we intend to measure? For our particular case, this was the (relative) extroversion of the children who took part in the study: the issue is whether the questionnaire used was adequate given the context in which it was completed, and whether the teacher-ratings (using the relative rating method, figure 1)

can give a 'true' reflection of the children's extroversion in the context of the classroom at least – and indeed whether this could be different from a rating in a different context. Validation of the questionnaire [4] suggests that it is a reliable measure, but does this extend across all contexts? We assumed that rating extroversion would be reasonably assessed by the teachers given that it relates to observable behaviours (as well as attitudes) that would be reasonably expected to be manifested by the children in the classroom environment.

IV. Conclusions

In describing these apparently problematic elements of the present study (section III), we do not seek to suggest that there is no value in exploring personality characteristics and its relationship to behaviours and performance. Instead, the case study presented here suggests that the methodological and reporting standards of such characteristics require clarity, in line with similar suggestions for the field of HRI in general [13]. In order to facilitate this, and to promote generalisation to (and comparison with) other studies, we suggest the following (modest) guidelines:

G1. Identify the source of the personality questionnaire (or other characterisation method) used in terms of the assumed dimensions (e.g. the 'Big 5' or the Eysenck dimensions), and whether it has been validated with the age group (and indeed language) under consideration.

G2. Identify the context in which the children completed the personality characterisation, and indicate possible influencing factors (e.g. completed in the presence of teachers/friends/parents, at home/school, in group/individually).

These guidelines are not particularly novel, and do in fact simply promote the complete reporting of measures and possible confounds. However, through our exploratory study we hope to have demonstrated that an apparent straightforward characterisation of one aspect of personality involves a number of complicating factors that should themselves be characterised. If the results we obtained were anomalous in some way, we hope that by reporting these potential confounds other researchers can build on them, by either accounting for the effect, or discounting it through further investigation. At the present time however, in our discussion of the results we highlight the possibility that child self-ratings of extroversion may be unreliable for child-robot interaction studies, whether this is due to inherent age-related unreliability, environment context effects, or others.

We do not suggest that we have a solution to the apparent issues described in this paper, particularly the mismatch between child- and teacher-ratings of extroversion, although we do venture some ideas for why this occurred. Indeed, we recognise a number of limitations in the study that prevent the formulation of a solution. Nevertheless, the suggestion remains that the application of personality assessments in a child-robot interaction context may be more complex than may be initially envisaged, with some dependence on context. As such, we suggest that the proposed guidelines will at least provide a basis upon which progress can be made.

References

[1] A. Tapus, C. Tapus, and M. J. Matarić, "User–robot personality matching and assistive robot behavior adaptation for post-stroke rehabilitation therapy," *Intelligent Service Robotics*, vol. 1, no. 2, pp. 169–183, 2008.

[2] K. Abe, et al, "Toward playmate robots that can play with children considering personality," *Proceedings of the second international conference on Human-agent interaction - HAI '14*, pp. 165–168, 2014.

[3] T. Belpaeme, et al, "Multimodal Child-Robot Interaction: Building Social Bonds," *Journal of Human-Robot Interaction*, vol. 1, no. 2, pp. 33–53, 2012.

[4] C. Barbaranelli, G. V. Caprara, A. Rabasca, and C. Pastorelli, "A questionnaire for measuring the Big Five in late childhood," *Personality and Individual Differences*, vol. 34, no. 4, pp. 645–664, 2003.

[5] L. J. Francis, "The development of an abbreviated form of the revised junior eysenck personality questionnaire (jepqr-a) among 13–15 year olds," *Personality and Individual Differences*, vol. 21, no. 6, pp. 835–844, 1996.

[6] L. J. Francis and E. M. Thomas, "Welsh language adaptation of the short-form junior eysenck personality questionnaire revised (jepqr-s)," *Psychologist in Wales*, vol. 21, pp. 25–32, 2008.

[7] J. Riding and V. Dyer, "The relationship between extraversion and verbal-imagery learning style in twelve-year-old children," *Personality and Individual Differences*, vol. 1, no. 3, pp. 273–279, 1980.

[8] T. Bidjerano and D. Y. Dai, "The relationship between the big-five model of personality and self-regulated learning strategies," *Learning and Individual Differences*, vol. 17, no. 1, pp. 69–81, 2007.

[9] P. Baxter, et al, "The Wider Supportive Role of Social Robots in the Classroom for Teachers," in *1st Int. Workshop on Educational Robotics at the Int. Conf. Social Robotics*, Paris, France, 2015.

[10] S. M. Robben, "It's NAO or Never Facilitate Bonding Between a Child and a Social Robot: Exploring the Possibility of a Robot Adaptive to Personality ." Radboud Universiteit, Nijmegen, The Netherlands, Tech. Rep., 2011.

[11] D. M. Buss, J. H. Block, and J. Block, "Preschool activity level: Personality correlates and developmental implications," *Child Development*, vol. 51, no. 2, pp. 401–408, 1980.

[12] I. Baroni, et al, "What a robotic companion could do for a diabetic child," in *The 23rd IEEE International Symposium on Robot and Human Interactive Communication (RoMAN'14)*. Edinburgh, U.K.: IEEE Press, 2014, pp. 936–941.

[13] P. Baxter, et al, "From Characterising Three Years of HRI to Methodology and Reporting Recommendations," in *alt.HRI at 11th Int. Conf. on Human-Robot Interaction*, vol. in press. Christchurch, NZ: ACM/IEEE, 2016.

Appendix: Adapted Extroversion Questionnaire

The adapted child-personality questionnaire (BFQ-C; Extroversion scale) used is as shown below [4]. Each Likert scale question had 3 possible responses: [*Almost Never, Sometimes, Almost Always*]. Answers were scored from 1 to 3, respectively, with all responses scored positively. Maximum range of possible responses: [*13, 39*]. Actual range of responses recorded: [*21, 37*], m=31, sd=3.137, n=38.

Q1) I like to meet with other people.
Q2) I like to compete with others.
Q3) I like to move and to do a great deal of activity
Q4) I like to be with others.
Q5) I can easily say to others what I think.
Q6) I say what I think.
Q7) I do something not to get bored.
Q8) I like to talk with others.
Q9) I am able to convince someone of what I think.
Q10) When I speak, the others listen to me and do what I say.
Q11) I like to joke.
Q12) I easily make friends.
Q13) I am happy and lively.