Robots for Future Classrooms: A Cross-Cultural Validation Study of "Negative Attitudes Toward Robots Scale" in the U.S. Context

Yu Xia[1] · Gerald LeTendre[1]

Abstract

Social robots are increasingly used in educational settings, but research on human–robot interaction in schools has largely focused on robotic functions and effects. Teacher attitudes towards robots are an important factor in determining the effectiveness of robot-related classroom interventions. In this paper we assessed the usefulness of the Negative Attitudes Toward Robots Scale (NARS) for use in U.S. populations. We reviewed research on the use of NARS and on teacher attitudes toward social robots, showing the potential of this scale for use in the field of education. Using data from 54 undergraduate students in a pre-service teacher education course, we found slightly different subscale loadings on the NARS than previous studies. We identified items on the scale that needed clarification and noted that the paucity of items related to emotional responses may limit the effectiveness of the scale as an international tool for assessing attitudes given cultural differences in emotion. We also found that lack of prior experience with robots was the strongest predictor of negative attitudes towards robots, suggesting that increased exposure to social robots in teacher education might be an effective way to improve educator attitudes toward robots.

1 Introduction

Technological advances have allowed the increased use of social robots, devices that can autonomously interact with humans in social settings [10, 30]. These robots are being studied for use in healthcare and educational settings [5, 16], and research has documented the wide range of factors that affect successful human–robot interactions (HRI) [42]. Work on social robots for use in education or health care can be situated in a broader field of work on socially intelligent computer/robotic systems [18], but the physical presence and attributes of social robots have a distinct effect from interactive learning programs accessed on a computer [12]. Previous research identified salient factors such as the social interfaces and features of robots [8, 48, 58], multimodal features of agentic robots [6], and also human participants' culture, gender, interest, and prior experiences with robots [2, 34, 49].

Social robots come in a wide variety of shapes and forms, see Belpaeme et al. [5] for a list of some of the most commonly deployed robots. Causo et al. [13] provide a summary and graphic depiction of many of the social robots currently being studied. The role that the robot is to play will be affected by its form and interactive capacities. For example, Causo et al. [13] note three broad forms of social robots—humanoid, semi-humanoid and pet-like. Belpaeme et al. [5] show that different robotic forms may be better suited for certain age ranges of students. While it is critical for social robots to have appropriate forms, functions, and be able to be socially responsive, integrating social robots into classroom settings also requires the active cooperation and engagement of classroom teachers, who may have significant concerns about robots [55].

While some previous research suggests that robotic features, rather than human factors, play the greatest role in affecting trust in HRI [24], other studies document the impact of human perceptions of robots [60]. We note previous studies that explored the attitudes of older individuals [35] and young children [61] with social robots, as well as a growing literature on teacher attitudes toward social robots [51, 55]. Because teachers are key gatekeepers in educational

✉ Gerald LeTendre
letendre@psu.edu

1 The Pennsylvania State University, 310G Rackley, University Park, PA 16802, USA

reform [17], the successful integration of social robots into K-12 classrooms will require the active participation of teachers. Yet teachers may have significant concerns about robots [55]. In education, positioning the human perspective in the classroom as a larger socio-technical learning system requires us to escape the microscopic approach to HRI and treat both humans and robots as agentic entities. Thus, on the human side, factors like gender, prior experience, and culture can all play role in affecting teacher attitudes toward robots. The investigation of social robots in educational settings requires research on teacher attitudes and concerns [13] and research on the role that human factors (e.g. gender, prior experience, and culture) may have as predictors of attitudes toward educational social robots. Our study focuses specifically on pre-service teacher attitudes and perceptions. In our review of the literature we analyze previous studies that focus on human reactions to social robots in education and provide a comprehensive review of studies investigating negative attitudes toward robots. We then analyze data on attitudes toward robots that were collected from likely future teachers to assess how they perceive social robots in general, and whether their gender, prior experience with interacting with robots, and their national background are associated with these attitudes.

2 Teachers and Social Robots in Education

Reviews of research on social robots in education document the rapid rise in interest in these devices as educational tools [4, 5, 39]. These reviews further document the effectiveness of social robots in working with students with autism disorders [4, 11] as well as in second-language acquisition [62]. However, comparatively few studies have examined human attitudes toward social robots in schools or issues arising from the long-term deployment of robots in real-life educational settings. As Benitti [7] wrote: "Most of the experiments involving robotics activities were not integrated into classroom activities, i.e., they occur as an after-school program or summer camp program" (p. 982).

If we are to understand how robots function in the wild [54], it is critical to understand the issues affecting their deployment and usage. Westlund et al. [63] provide a detailed guide for researchers intending to test social robots in schools, and they note: "Understanding the teachers, the classroom environment, and the constraints involved is especially important for microgenetic and longitudinal studies…" (p. 390). Significantly, they identify teachers as one of the key stakeholders to contact before initiating studies and advise future researchers working with social robots in classrooms to involve teachers early on and pay close attention to their concerns. Teacher attitudes, experience and beliefs are salient factors which are likely to affect the effective

deployment of robots. Some research has even suggested that there are significant school-to-school differences in staff attitudes toward social robots [31].

Research on teacher experiences with robots has discovered a mix of positive and negative attitudes. Majgaard [37] found that teachers thought that robots motivated students and were an excellent "object to think with" (pp. 80-81). However, other studies also identified teacher concerns about disruption, inclusion and time needed to integrate robots into the classroom [55, 56, 63]. Causo et al. [13] found that "The teachers experienced technical issues that, while managed in a variety of ways, could discourage mass-scale adoption" (p. 4269). Chang, Lee, Chao, Wang, and Chen [14, p. 20] identified three overarching concerns that teachers initially held: anxiety that the robots would be damaged; that without support, the robots would not provide any instructional advantage; that robots would be too complicated to use effectively. Exploratory, qualitative work [55] similarly found that teachers had clear concerns about administrative burden and equitable access to the robot among children.

In addition to concerns about robots, teachers' lack of knowledge about social robots is also likely to be a major factor affecting whether educators are willing to consider deploying a robot in the classroom. In this regard, Jones et al. [27, p. 288] found that in addition to concerns about technical problems, teachers found it difficult to envision how a social robot could function in the classroom. He argued that teachers may find it difficult to estimate the full impact that new technologies might have on the classroom. As Hattie [26] has argued, if teachers do not see the effect of a new educational innovation, they will be unlikely to adopt it. This line of reasoning is supported by Fridin and Belokopytov [22, p. 29] who found that "…the desire to use robotic teacher assistants by preschool and elementary school teachers is determined mostly by their beliefs that the robotic assistant will enhance and facilitate the educational process." Diep, Cabibihan and Wolbring [19] noted that special education teachers resisted the adoption of social robots because they believed the robots could not communicate or interact emotionally with students. And, while Reich-Stiebert and Eyssel [51] did not find that gender, age or subject area affected teacher attitudes, they did find that "…teachers who reported higher interest in technology were more inclined to use education robots than teachers who were less interested in technological issues….". Finally, Kennedy et al. [31] argued that lack of interest in technology can present persistent barriers to implementation of robots in the classroom.

These studies suggest that teachers' initial attitudes towards robots, their knowledge about robots, and their interest in technology are important factors that appear to influence teacher willingness to work with social robots. Some studies show that pre-service teachers have similar

wide-ranging concerns about robots, particularly along moral and ethical lines [56], but it was not clear if these attitudes would inhibit their willingness to adopt social robots in the classroom. As Ahmad, Mubin and Orlando [1] showed, exposure to actual social robots appears to focus teacher concerns about robots on more specific tasks and functionalities. Lee and Kim [32] noted a significant reduction in negative attitude (as measured by the NARS – Negative Attitudes Toward Robots Scale) among pre-service teachers who took robotic programming. This suggests that some kind of pre-service intervention or exposure to social robots might affect teacher attitudes. Negative attitudes among pre-service teachers or teachers without significant interest or confidence in STEM education may be amendable to change given exposure to robots. Syrdal, Dautenhahn, Koay and Walters [57] found the NARS to be effective in differentiating attitudes in a series of HRI trials. Thus, the NARS may provide a simple and robust tool to effectively assess student teachers' attitudes toward social robots and might serve as an important tool for researchers, administrators and reformers working on introducing social robots in real-life classrooms.

3 Negative Attitudes Towards Robots Scale (NARS)

The NARS was developed by Nomura et al. [40–42] in Japan to measure general human attitudes toward robots. When developing the instrument, Nomura and his colleagues adopted a bottom-up approach and then modified their existing instruments. In their initial pilot study in 2003, they collected free writing answers using open-ended questionnaires from researchers and students in engineering and psychology programs. They developed a questionnaire using a sample of 265 Japanese university students and then validated it on a new sample of 240 Japanese university students [40]. They then extracted themes related to emotions and attitudes [45] from this data. They also modified two other scales (see 45) for more details), and after verifying internal consistency and validity, they selected 14 items out of the original 32 items they had considered.

The current NARS scale consists of 14 survey items. Nomura et al. [41] used exploratory factor analysis to classify these items into three subscales. S1 (six items) measures negative attitudes toward interactions with robots, S2 (five items) measures negative attitudes toward the social influence of robots, and S3 (three items) measures negative attitudes toward emotional interactions with robots. All 14 items are scored on a five-point scale: Strongly Disagree (1), Disagree (2), Neutral (3), Agree (4), Strongly Agree (5). Total scores are calculated by adding up individual scores after reverse coding items 12, 13, and 14.

The NARS has been widely used and its reliability and construct validity have been established in multiple context (for example 21, 60; see more in Table 1). In subsequent studies, researchers have adapted the NARS to new linguistic and cultural contexts [2, 57], thus opening the door to a larger research program on the cross-cultural study of human attitudes toward robots. However, while the NARS appears robust, as Syrdal et al. [57] noted, "the Negative Attitudes Towards Robots Scale may be susceptible to cultural differences." Within this vein of research, we note that since

Table 1 Selected studies that used NARS as measuring instrument

Study	Year	n	Language	Description
[40, 41]	2004 2006	263 240	Japanese	Development and validation of NARS
[2]	2005	96	English	Cross-cultural study on attitudes toward robots
[45]	2006	53	Japanese	Investigated the influence of negative attitudes toward robots on human–robot interaction
[46]	2006	400	Japanese	Investigated relationships between negative attitudes toward and assumptions about robots such as their types and tasks
[3]	2007	467	English	Investigated the influence of culture and prior experience on attitudes toward robots
[47]	2007	17	Japanese	Investigated relationships between negative attitudes and anxiety toward, and allowable distance of a robot
[43]	2008	38	Japanese	Investigated the relationships between people's attitudes and emotions, and their behavior toward a robot
[15]	2009	119	English	Investigated the relationships between people's attitude and their behaviors with robots including touch and proactivity
[57]	2009	28	English	Described the use of the NARS to explain participants' evaluations of robot behavior
[60]	2011	120	English	Reviewed the development and validation of NARS scale and reported three studies that used NARS
[44]	2015	200	Japanese English	Using NARS to establish criterion-validity for a new instrument called Frankenstein Syndrome Questionnaire
[21]	2018	210	English	Focus on the students' attitudes toward robots; analyzed the results and studied the factors that affect the perception of students toward robots
[29]	2018	24	English	Investigated the relationship between individual differences in attitudes toward robots and language learning

Serholt's paper in 2014, there is no study that investigates the attitudes toward robots for its use in educational contexts by pre-service teachers in the U.S.

If the NARS is to be used to assess pre-service teacher attitudes toward robots, research on this sub-population is needed. Reich-Stiebert and Eyssel [51, p. 674] hypothesized that age, gender and previous experience with robots would affect teacher attitudes toward robots. They found that factors including gender, age, teaching domain, and previous experience with robots did not affect attitudes toward robots, but they also noted that this lack of association may be due to sampling issues as there were more females than males in their sample and most respondents had no prior experience with robots. Fernandez-Llamas et al. [21] did not analyze gender but did find differences by age. Given the limited age range of students in our sample, we assume there will be no significant association with age. In line with Reich-Stiebert and Eyssel [51], we anticipate that males will have more positive attitudes toward robots than females, and that those with more previous experience will have more positive attitudes than those with less experience. While Lee and Sabanovi [33] showed overall low evaluations of robots as companions or teachers, independent of the cultural origin (e.g. USA, Turkey, and South Korea), following Kamida, Hiroko and Arai [28] we expect that U.S. students will have more positive attitudes than international students. To address our research goal and test the hypotheses, we ask the following research questions:

1. Did the factor analysis render same factor structure for the US population as compared to validation studies with populations from other cultural backgrounds?
2. Were there significant differences negative attitudes towards robots by gender, national background, and prior experience with robots?

4 Methods

Previous research indicates that teachers have a wide range of concerns about deploying social robots in the classroom. It is clear then, that the successful implementation of robots must take into account teacher's concerns and knowledge about robots, not just student engagement with these devices [52]. Preservice teachers represent a population that typically have little exposure to robots. We wished to assess how students in an undergraduate pre-service class would respond to the idea of robots in the classroom. Through assessing the usefulness of the NARS with preservice teachers, we can identify a simple, robust scale that administrators, reformers or teacher educators can effectively use to measure attitudes toward robots. This could then lead to better assessment of what interventions might be needed

to prepare future teachers for working with social robots in the classroom.

4.1 Sample and Participants

The sample consists of 58 undergraduate students enrolled in an introductory education course dealing with technology and learning in a large university in the Mid-Atlantic region of the United States. Most of the students intend to pursue some form of education degree, although at this point in their studies, some of the students may not have declared a major. The students in this sample are in the earliest stage of pre-service teacher education, one where students are still considering their commitment to a teaching major. Participation in the study was voluntary; respondents were not compensated. As part of the course, students were introduced to Pepper (see 13), a social robot currently available for commercial purchase and use in classrooms around the world. The students observed the robot operating under its "autonomous life" programming that acts as a default system for basic social interactions. We chose not to modify the programming so as to emulate the conditions a classroom teacher (with no programming training) would face unpacking and operating a social robot for the first time. In its "autonomous life" mode, Pepper is capable of answering simple questions and performing simple dances.

The sample size was not determined in advance. Based on reviewer comments to the first version of the paper, we dropped all non-NARS items for this analysis. Four participants who answered only the first four questions only and did not give scores to any of the items in the scale, were removed from the final data set leaving 54 students. Among the 54 participants, the majority are between the age from 18 to 25 (96.30%, $N=52$); there are 13 males (24.07%) and 41 females (75.93%).

4.2 Data Collection

After the module on Innovation and Culture, where students had the chance to interact with Pepper, they were sent a link to a website and invited to take part in a survey that included the NARS items. Data was collected using an online survey. The survey started with a question asking consent to participate in the survey titled Social Robots Survey hosted in an online survey website (https://pennstate.qualtrics.com), followed by open-ended questions asking age and gender, a multiple choice on national background (U.S. or International) and prior experience with robot (None, Some, A Lot). Participants answered a total of 39 questions about their attitudes toward robots including 14 questions replicating the Negative Attitudes Toward Robots Scale (NARS) developed by Nomura et al. [40–42].

4.3 Data Analysis

In the current study, we analyzed data that asked about participants' age, gender, national background (U.S. or International), and prior experience with robots, and scores on the five-point Likert questions in the NARS. Using SPSS Statistics 26, we first studied the reliability of the scores, before conducting confirmatory factor analysis. To do so, items 12, 13, and 14 from NARS scale were reverse scored. There was 1 missing value in items 4, 5, 7, and 8, which was replaced with the series means.

We then conducted correlation analyses to examine internal reliability for the scale, as well as the three subscales. Aiming to establish the validity with a new population, we conducted confirmatory factor analysis using the R lavaan package (http://lavaan.ugent.be/) [53]. While the goodness of fit indices showed acceptable results, we still conducted an exploratory factor analysis to check for the possibility of a different factor structure. This potential factor structure was then compared to the original structure to yield insights into the differences between the two populations, e.g. our sample and the sample used by Nomura. In considering the extraction method for exploratory factor analysis, since we were attempting to understand the latent factors among items [64], we chose a common factor analysis method, principal axis factoring.

To further investigate the relationship between gender, national background, prior experience with robots, and participants' negative attitudes towards robots, three two-way ANOVA tests were first conducted for the total scores from NARS scale, and additional three two-way ANOVA analyses were conducted for the three subscale total scores. We further conducted a post hoc and a priori power analyses with the program *G*Power* [20].

5 Results

5.1 Preliminary Analyses for Reliability and Validity

After collecting and cleaning the survey scores from the undergraduate students, we replaced missing values and reversed coded three items. The Shapiro–Wilk test of normality was $p = .162$, showing that there was no statistically significant violation of the normality assumption. An initial reliability analysis using Cronbach's Alpha ($\alpha = .88$) was conducted. While Cronbach's Alpha would increase if two items were deleted (These items are: "2. The word "robot" means nothing to me" and "11. I feel that in the future society will be dominated by robots"), we decided to keep the full scale with 14 items because the Cronbach's Alpha of .88 already supports the consistency among scores in measuring a unidimensional construct. As the scale is

composed of three subscales with S1 measuring negative attitudes toward interactions with robots, S2 negative attitudes toward the social influence of robots, and S3 negative attitudes toward emotional interactions with robots, further correlation analyses were conducted to examine the internal consistency within subscales, yielding Cronbach's Alphas of .818 (S1), .773 (S2), and .761 (S3) respectively. These Alphas range from acceptable to good according to George and Mallery [23].

A confirmatory factor analysis was conducted using the R lavaan package to test the structural validity of NARS for data fit. The evaluation of data fit under the CFA in this study is based on a commonly used Chi square test statistic in combination with several other goodness of fit indices. While the Chi square value is statistically significant ($\chi^2_{62} = 84.17$, $p = .03$), the evaluation of data fit was based on a joint examination of other indices including *the comparative fit index* (CFI), *the Tucker–Lewis index* (TLI), *the root mean square error of approximation* (RMSEA) with its 90% confidence interval (CI), and *the standardized root mean square residual* (SRMR). The evaluation is often determined by satisfactory thresholds with RMSEA ≤ 0.06, TLI ≥ 0.95, CFI ≥ 0.95), SRMR ≤ 0.08 [9]. However, Marsh, Hau, and Wen [38] also recommended less strict criteria of a reasonable fit in practical applications with CFI $\geq .90$, TLI $\geq .90$, and RMSEA $\leq .08$. The confirmatory factor analysis on our sample data yielded a CFI of .93, TLI of .91, RMSEA of .08, and SRMR of .08, which just meet the reasonable thresholds. These indices warrant an exploratory approach to investigate the differences between scores obtained from our population of US preservice teachers and those obtained from undergraduate students in Japan by Nomura et al. [40–42].

Considering the small sample size in our study, we conducted KMO and Barlett's tests to see if we had sampling adequacy. The results indicated that we met the minimum standard for factor analysis with KMO = .807 and $\chi^2_{62} = 357.08$, $p < .001$ for Bartlett's test of sphericity. We continued with principal axis factoring with varimax rotation. Table 2 shows the solution from this factor analysis, with an added column of original subscales from Nomura et al. [45] for better comparison. Factor loadings lower than .40 that did not meet the Harman's criterion [25] were deleted from the table except for item 11. Overall, our analysis yielded four subscales rather than three. In terms of specific items, items 1 and 12 showed close cross loadings across factors; item 7 also had cross loadings across factors 1 and 3 but the loading on factor 1 was much higher than factor 3. Item 11 was the most problematic with low loadings on all factors; the highest being .253 on factor 1.

A comparison with the original subscale solution shows that our factors 1 and 3 were consistent with S2 and S3 from Nomura et al. [45], but items in Nomura et al.'s subscale 1 were loaded among factors 1, 2, and 4. Remember

Table 2 Subscale factor loadings from principal axis factoring with Varimax rotation

Original sub-scales	Items	Subscales			
		1	2	3	4
S1	1. I would feel uneasy if I was given a job where I had to use robots	*.570*	.481		
S1	2. The word "robot" means nothing to me				***.435***
S1	3. I would feel nervous operating a robot in front of other people				***.542***
S1	4. I would feel very nervous just standing in front of a robot		***.702***		
S1	5. I would hate the idea that robots or artificial intelligences were making judgements about things		***.710***		
S1	6. I would feel paranoid talking with a robot		***.672***		
S2	7. I would feel uneasy if robots really had emotions	**.713**		.431	
S2	8. Something bad might happen if robots developed into living beings	**.753**			
S2	9. I feel that if I depend on robots too much, something bad might happen	**.661**			
S2	10. I am concerned that robot would be a bad influence on children	**.601**			
S2	11. I feel that in the future society will be dominated by robots	**.253**			
S3	12. I would feel relaxed talking with robots*			**.599**	.419
S3	13. If robots had emotions, I would be able to make friends with them*			**.752**	
S3	14. I feel comforted being with robots that have emotions*			**.727**	

Bold numbers are the highest factor loadings

Bold-italic numbers are originally on S1 in Nomura et al. [45] but loaded on different factors in our analysis

*Reverse coded items

that S1 measures negative attitudes toward interactions with robots, S2 measures negative attitudes toward the social influence of robots, and S3 measures negative attitudes toward emotional interactions with robots. That is, in our sample student perceptions of item 1 indicate it was more associated with the social influence of robots rather than interactions with robots. However, the finding is ambiguous as the loading on factor 1 (.570) is only slightly higher than the loading on factor 2 (.481). Item 7 was crossed loaded on factor 3, also the S3 from Nomura et al., indicating that the word "emotions" might contribute to the cross loading on negative attitudes toward emotional interactions with robots. Interestingly, a new factor emerged that includes items 2 and 3 and potentially item 12.

5.2 NARS Differences in Terms of Gender, National Background, and Prior Experience

To see how the NARS scores related to the three factors of gender (male vs. female), national background (US vs. international), and prior experience for interaction with robots (none vs. some vs. a lot), we conducted three two-way ANOVA tests. Average NARS scores for each participant, which could range from 0 to 60, and the means as a function of gender, national background, and experience are shown in Table 3. The F-test results for main effects and interactions of the three two-way ANOVA tests are shown in Tables 4, 5 and 6.

There were no statistically interactions between all three pairs of independent variables: gender × national background ($p = .334$), national background × experience

Table 3 Mean NARS scores (standard deviation) across gender, national background, and experience levels

Gender	National background	Experience			
		None ($n=34$)	Some ($n=18$)	A Lot ($n=2$)	Total
Male ($n=13$)	US ($n=8$)	56.00	35.14 (6.91)		37.75 (9.77)
	Int. ($n=5$)	36.77 (7.55)	38.00	21.00	33.86 (8.97)
	Total	41.58 (11.42)	35.50 (6.48)	21.00	36.26 (9.29)
Female ($n=41$)	US ($n=37$)	47.06 (6.81)	47.50 (3.89)	27.00	46.61 (6.99)
	Int. ($n=4$)	37.50 (19.09)	35.50 (16.26)		36.50 (14.53)
	Total	46.42 (7.85)	45.10 (8.17)	27.00	45.63 (8.30)
Total		45.85 (8.28)	40.83 (8.76)	24.00 (4.24)	43.37 (9.38)

Table 4 Two-way ANOVA for NARS (gender × national background)

Source	Sum of Squares	df	F	p value	η_p^2	Observed power[b]
Corrected model	1282.51[a]	9	6.33	.001	.28	.95
Error	3379.13	50				
C. Total	4661.64	53				
Gender	219.65	1	3.25	.077	.06	.42
National background	325.54	1	4.82	.033*	.09	.58
Gender × national background	64.42	1	.95	.334	.02	.16

*p < .05

[a]R Squared = .275 (adjusted R squared = .232)

[b]Computed using alpha = .05

Table 5 Two-way ANOVA for NARS (national background × experience)

Source	Sum of Squares	df	F	p value	η_p^2	Observed Power[b]
Corrected model	1619.68[a]	5	5.11	.001	.35	.97
Error	3041.96	48				
Corrected total	4661.64	53				
National background	178.82	1	2.82	.099	.06	.38
Experience	604.64	2	4.77	.013*	.17	.77
National background × experience	40.81	2	.32	.726	.01	.10

*p < .05

[a]R squared = .347 (adjusted R squared = .279)

[b]Computed using alpha = .05

Table 6 Two-way ANOVA for NARS (gender × experience)

Source	Sum of Squares	df	F	p value	η_p^2	Observed Power[b]
Corrected model	1586.31[a]	5	4.95	.001	.34	.97
Error	3075.34	48				
Corrected total	4661.64	53				
Gender	166.61	1	2.60	.113	.05	.35
Experience	713.43	2	5.57	.007**	.19	.83
Gender × experience	45.57	2	.356	.703	.02	.10

*p < .01

[a]R squared = .340 (adjusted R squared = .272)

[b]Computed using alpha = .05

with interaction with robots ($p = .726$), and gender × experience with interaction with robots ($p = .703$). In the results of two-way ANOVA of gender × national background, a significant difference was found between national backgrounds, $F(1,50) = 4.82$, $p = .033$, $\eta_p^2 = .09$ at the .05 level. We further conducted a post hoc power analysis with the program *G*Power* [20]. Using partial eta-squared, the Cohen's f was calculated to be .31, and the power was .62 at the .05 level. A priori power analysis indicated that a sample size of 82 would be needed to detect a significant main effect by national background with a power of .80 and an alpha of .05.

The two-way ANOVA of national background × experience showed a significant difference cross experience levels, $F(2,48) = 4.77$, $p = .013$, $\eta_p^2 = .17$ at the .05 level. Post hoc analysis (Scheffe) revealed significant differences between a lot and none experience ($p = .002$), and between a lot and some experience ($p = .024$). Students without any experience ($M = 45.85$, $SD = 8.28$) had a higher negative attitude score than students with a lot experience ($M = 24.00$, $SD = 4.24$),

and students with some experience ($M=40.83$, $SD=8.76$) had a higher negative attitude score than students with a lot experience ($M=24.00$, $SD=4.24$). For national background, students from the US ($M=45.04$, $SD=8.18$) scored much higher than those from international background ($M=35.04$, $SD=11.01$). We again conducted a post hoc power analysis with the program *G*Power*. Using partial eta-squared, the Cohen's *f* was calculated to be .45, and the power was .90 at the .05 level.

The two-way ANOVA of national background × experience showed a significant difference cross experience levels, $F(2,48)=5.57$, $p=.007$, $\eta_p^2=.19$ at the .05 level. A post hoc power analysis with the program *G*Power* was conducted. Using partial eta-squared, the Cohen's *f* was calculated to be .48, and the power was .94 at the .05 level. While no significant difference was found for gender, NARS scores for females ($M=45.63$, $SD=8.30$) were found to be higher than those for males ($M=36.26$, $SD=9.29$), indicating that female students held more negative attitudes toward social robots than male students.

Given a reasonable model data fit from the confirmatory factor analysis, we continued the statistical analyses on the three subscale total scores to see whether or not there were significant differences by gender, national background, and previous experience with robots.

For S1—negative attitudes toward interactions with robots—no statistically significant interactions were found for all interactions, $p=.419$ between gender and national background, $p=.269$ between gender and experience, $p=.256$ between national background and experience, and $p=.277$ between all the three independent variables at the .05 level. In terms of main effects, no significant difference was found for gender. However, there was a significant difference for national background, $F(1,44)=5.77$, $p=.021$, $\eta_p^2=.12$ at the .05 level. Students from the US ($M=17.58$, $SD=4.72$) scored much higher than those from international backgrounds ($M=12.75$, $SD=3.59$). There was a significant difference for experience level, $F(2,44)=5.20$, $p=.009$, $\eta_p^2=.19$. The post hoc analysis (Scheffe) result was consistent with the results for the NARS total score; again, students without any experience ($M=18.17$, $SD=4.46$) had a higher negative attitude score than students with a lot experience ($M=8.00$, $SD=0.00$). Further, students with some experience ($M=15.11$, $SD=4.40$) had a higher score than students with a lot experience ($M=8.00$, $SD=0.00$).

For S2—negative attitudes toward the social influence of robots—no statistically significant differences were found for any of the interactions, $p=.857$ between gender and national background, $p=.839$ between gender and experience, $p=.185$ between national background and experience, and $p=.077$ between all the three independent variables; or main effects, $p=.751$ for gender, $p=.166$ for national background, and $p=.074$ for experience level.

For S3—negative attitudes toward emotional interactions with robots—again no statistically significant differences were found for interactions, $p=.568$ between gender and national background, $p=.057$ between gender and experience, $p=.704$ between national background and experience, and $p=.374$ between all the three independent variables. While there was no significant difference by gender, $p=.749$, there were significant differences by national background ($F(1,44)=5.67$, $p=.022$, $\eta_p^2=.11$) and by experience level ($F(2,44)=5.03$, $p=.011$, $\eta_p^2=.19$). The post hoc analysis (Scheffe) result revealed that again, students without any experience ($M=10.74$, $SD=2.27$) had a higher negative attitude score than students with a lot experience ($M=6.00$, $SD=1.41$). Further, students with some experience ($M=9.83$, $SD=2.26$) had a higher score than students with a lot experience ($M=6.00$, $SD=1.41$).

6 Discussion

This study focuses on teacher attitudes and perceptions towards social robots by implementing the NARS scale [40–42] with a sample of undergraduates in a teacher preparation course. The scale was developed and validated to assess three latent constructs: negative attitudes toward actual interactions with robots (S1), negative attitudes toward the social influence of robots (S2), and negative attitudes toward emotional interactions with robots (S3). The exploratory factor analysis helped us see the differences in constructs between this US population and a Japanese population of undergraduates. The comparison provides insights that could lead to future improvement of certain items. For example, subscales 2 and 3 from the original studies had identical items as in our study (Table 2), but subscale 1's items did not load on one factor.

To improve the NARS for use with U.S. teacher and pre-service teacher populations, we suggest the following changes. First, item 1 needs to be rephrased. In Nomura et al. [45], it was associated with actual interaction with robots, whereas in this population, it is associated with the social influence of robots. That is, university students from Japan might have focused more on "I had to use robots," i.e. the actual interaction, whereas participants in our study might have put more emphasis on "if I was given a job," thus creating an association with social influence.

Regarding item 2, since deleting the item would improve the overall reliability from .877 to .886, and the subscale reliability from .818 to .846, this item may need to be replaced or reworded. We would also suggest replacing or rewording item 11 since deleting the item would improve the overall reliability from .877 to .885, and the subscale reliability from .773 to .834. Interestingly, a previous cross-cultural study that used the NARS on a sample from

a British university showed similar problems for items 2 and 11, which, along with item 3, were deleted after a reliability analysis on the overall scale. The fact that items 2 and 11 were problematic for participants from both British and American universities suggests that the current English phrasing remains an issue.

Item 2 was originally categorized into the first factor, negative attitudes toward situations of interaction with robots [45]. However, reading from the face validity, the wording of "The word 'robot' means nothing to me" seems too general and does not necessarily speak to any "situations of interaction with robots." Item 11 was originally grouped into the second subscale of negative attitudes toward the social influence of robots. The face validity of this item does imply social influence simply by describing a possible scenario in which robots dominate a future society. It may be that respondents found some ambiguity in the word "dominate" in terms of in what ways and to what extent robots would act. More extensive interviews or focus groups would be needed to allow us to understand how respondents interpreted this particular item.

As for the overall scale, we suggest adding one or two additional items about emotional interactions with robots since S3 only has three items. In factorial analysis, three or four indicators for one factor is considered very few, if not unacceptable [36]. Raubenheimer [50, p. 60] directly pointed out that "since the number of items per factor is crucial. Specifically……it would require at least four items". While the scale has continued to be widely used in the field, we suggest that an expansion of the scale would enable researchers to be in a better position to make claims about participants' attitudes toward robots. Since emotions are culturally conditioned, expanding this subscale could provide important information about the cultural effects at play on the human side of the HRI.

After conducting three sets of two-way ANOVAs on NARS total scores and additional three sets of three-way ANOVA analyses on subscale totals to identify differences based on gender, national background, or experience with robots, we found evidence that indicates the importance of national background and prior experience but not gender. The effect of national background and prior experience with robots on negative attitudes was evidenced in the significant differences in NARS total scores, S1, and S3. Although differences were found for both variables, experience appeared to have more practical significance given its larger effect size. Although gender was not a significant factor, female participants ($M = 36.50$, $SD = 14.53$) scored higher than male participants ($M = 36.26$, $SD = 9.29$) in our sample. This is consonant with Nomura et al. [46] who reported that compared to male students who participated in the study, female students had stronger negative attitudes toward interactions with robots but lower negative attitudes toward emotions in interaction with robots. There was no gender difference in regard to their negative attitudes toward the social influence of robots. However, another study by Nomura and his colleague [43] reported no effects for gender in the NARS scores. According to the National Center for Education Statistics, in 2015-2016, 77% of public school teachers were female, a slight increase from 75% in 2000 [59]. A similar percentage was seen in our sample where female students accounted for 75.9% of the total respondents. A more pronounced negative attitudes toward robots among female pre-service teachers would be important to note with regard to future integration of robots into classrooms.

7 Conclusion

In this study, we adopted a widely used instrument used in assessing negative attitudes towards robots called NARS [45] to investigate pre-service teachers' attitudes. A confirmatory factor analysis established a reasonable, but not perfect, goodness-of-fit evaluation. We thus conducted an exploratory factor analysis. Our results suggest that for U.S. pre-service teacher populations, some improvement of the NARS is needed, especially in regard to items 1, 2, and 11. Given that almost 200,000 students are enrolled in teacher education programs in the U.S., improvement of the NARS would provide the robotic and educational research communities with a robust tool for measuring attitudes toward social robots.

The results of this study have implications for both teacher education and professional development programs. While research has documented positive student learning gains with social robots as tutors, peer learners or storytellers in educational settings, we still do not yet know how best to prepare teachers to work with social robots. Given the increasing presence of robots in K-12 classrooms, getting to know the attitudes of pre-service teachers towards robots appears vital. If social robots were introduced into teacher education or professional development, the NARS could be an important tool in predicting how pre-service or in-service teachers would interact with robots, and could be used to study how their attitudes changed after interactions with robots. Considering that our study found that prior experience was an important factor affecting attitudes toward robots, it is likely that increased exposure to robots in teacher education and professional development may be necessary to promote successful HRI in the classroom.

The different factor loadings that we found in this population suggest that researchers may need to rethink the differentiation between constructs such as individual reactions to personal contacts with robots and the influence that robots are believed to have at the societal level. This may be a more crucial distinction for social robots, as opposed

to other forms of robots, as social interactivity requires an emotional component. NARS appears to be limited in regard to measuring emotional responses to robots. Some items in the current NARS versions may need to be worded in ways that decrease ambiguity for respondents. We suggest future research consider alternative items that measure emotional construct towards robots. Such research might also shed light on emotional differences in attitudes toward robots arising from cultural background. Given the cultural diversity in the U.S., it is possible that different groups may have different emotional reactions to robots.

Also, given the important role that prior experience with robots plays in participants' negative attitudes, there is a need to incorporate robots into teacher education or professional development programs. Interaction between children or learners and robots has been studied in its various forms, but we should not forget that interaction with robots will also be a new experience for the vast majority of teachers. Teachers, students, and social robots will all be key actors in future classrooms. These socio-technical learning environments will still be directed by teachers, and so more thought and consideration needs to be given to how teachers' attitudes and beliefs will influence the future deployment of social robots.

References

1. Ahmad MI, Mubin O, Orlando J (2016) Understanding behaviours and roles for social and adaptive robots in education: teacher's perspective. In: Proceedings of the fourth international conference on human agent interaction. ACM, pp 297–304
2. Bartneck C, Nomura T, Kanda T, Suzuki T, Kato K (2005) Cultural differences in attitudes towards robots. In: Proceedings of symposium on robot companions (SSAISB 2005 convention), pp 1–4
3. Bartneck C, Suzuki T, Kanda T, Nomura T (2007) The influence of people's culture and prior experiences with Aibo on their attitude towards robots. AI Soc 21(1–2):217–230
4. Begum M, Serna RW, Yanco HA (2016) Are robots ready to deliver autism interventions? A comprehensive review. Int J Soc Robot 8(2):157–181
5. Belpaeme T, Kennedy J, Ramachandran A, Scassellati B, Tanaka F (2018) Social robots for education: a review. Sci Robot 3:1–9
6. Belpaeme T, Baxter P, Read R, Wood R, Cuayáhuitl H, Kiefer B et al (2013) Multimodal child–robot interaction: building social bonds. J Hum Robot Interact 1(2):33–53
7. Benitti FBV (2012) Exploring the educational potential of robotics in schools: a systematic review. Comput Educ 58(3):978–988
8. Bernotat J, Eyssel F, Sachse J (2019) The (Fe) male robot: how robot body shape impacts first impressions and trust towards robots. Int J Soc Robot 11:1–13
9. Boateng GO, Neilands TB, Frongillo EA, Melgar-Quiñonez HR, Young SL (2018) Best practices for developing and validating scales for health, social, and behavioral research: a primer. Front Public Health 6:149
10. Breazeal C (2002) Designing sociable robots. MIT Press, Cambridge
11. Cabibihan JJ, Javed H, Ang M, Aljunied SM (2013) Why robots? A survey on the roles and benefits of social robots in the therapy of children with autism. Int J Soc Robot 5(4):593–618
12. Causo A, Vo GT, Chen I-M, Yeo SH (2016) Design of robots used as education companion and tutor. In: Zeghloul S (ed) Robotics and mechatronics, vol 37. Springer, Cham, pp 75–84
13. Causo A, Win PZ, Guo PS, Chen I-M (2017) Deploying social robots as teaching aid in pre-school K2 classes: a proof-of-concept study. Paper presented at the 2017 IEEE international conference on robotics and automation, Singapore
14. Chang CW, Lee JH, Chao PY, Wang CY, Chen GD (2010) Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. J Educ Technol Soc 13(2):13–24
15. Cramer H, Kemper N, Amin A, Wielinga B, Evers V (2009) 'Give me a hug': the effects of touch and autonomy on people's responses to embodied social agents. Comput Anim Virtual Worlds 20(2–3):437–445
16. Dahl T, Boulos MK (2014) Robots in health and social care: a complementary technology to home care and telehealthcare? Robotics 3(1):1–21
17. Darling-Hammond L (2010) The flat world and education. Teachers College Press, New York
18. Dautenhahn K, Bond AH, Canamero L, Edmonds B (eds) (2002) Socially intelligent agents: creating relationships with computers and robots (multiagent systems, artificial societies, and simulated organizations). Springer, Berlin
19. Diep L, Cabibihan JJ, Wolbring G (2015) Social robots: views of special education teachers. In: Proceedings of the 3rd 2015 workshop on ICTs for improving patients rehabilitation research techniques. ACM, pp 160–163
20. Erdfelder E, Faul F, Buchner A (1996) GPOWER: a general power analysis program. Behav Res Methods Instrum Comput 28(1):1–11
21. Fernandez-Llamas C, Conde MA, Rodríguez-Lera FJ, Rodríguez-Sedano FJ, García F (2018) May I teach you? Students' behavior when lectured by robotic vs. human teachers. Comput Hum Behav 80:460–469
22. Fridin M, Belokopytov M (2014) Acceptance of socially assistive humanoid robot by preschool and elementary school teachers. Comput Hum Behav 33:23–31
23. George D, Mallery P (2003) SPSS for Windows step by step: a simple guide and reference. 11.0 update, 4th edn. Allyn & Bacon, Boston
24. Hancock P, Billings D, Schaefer K, Chen J, Visser ED, Parasuraman R (2011) A meta-analysis of factors affecting trust in human–robot interaction. Hum Fact J Hum Fact Ergonom Soc 53:517–527
25. Harman HH (1976) Modern factor analysis, 3rd revised edn. University of Chicago Press, Chicago
26. Hattie J (2012) Visible learning for teachers. Routledge, New York
27. Jones A, Küster D, Basedow CA, Alves-Oliveira P, Serholt S, Hastie H et al (2015) Empathic robotic tutors for personalised learning: a multidisciplinary approach. In: International conference on social robotics. Springer, Cham, pp 285–295
28. Kamide H, Arai T (2017) Perceived Comfortableness of Anthropomorphized Robots in U.S. and Japan. Int J Soc Robot 9:537–543
29. Kanero J, Franko I, Oranç C, Uluşahin O, Koşkulu S, Adıgüzel Z, et al (2018) Who can benefit from robots? effects of individual differences in robot-assisted language learning. In: 2018 joint IEEE 8th international conference on development and learning and epigenetic robotics (ICDL-EpiRob). IEEE, pp 212–217
30. Kanda T, Hirano T, Eaton D, Ishiguro H (2004) Interactive robots as social partners and peer tutors for children: a field trial. Hum Comput Interact 19:61–84

31. Kennedy J, Lemaignan S, Belpaeme T (2016) The cautious attitude of teachers towards social robots in schools. In: Robots 4 learning workshop at IEEE RO-MAN 2016

32. Lee Y, Kim SW (2016) An exploration of attitudes toward robots of pre-service teachers' through robot programming education. In: E-learn: world conference on e-learning in corporate, government, healthcare, and higher education. Association for the Advancement of Computing in Education (AACE), pp 737–741

33. Lee HR, Sabanović S (2014) Culturally variable preferences for robot design and use in South Korea, Turkey, and the United States. In: Proceedings of the 2014 ACM/IEEE international conference on human–robot interaction. ACM, pp 17–24

34. Li D, Rau P, Li Y (2010) A cross-cultural study: effect of robot appearance and task. Int J Soc Robot 2:175–186

35. Louie WYG, McColl D, Nejat G (2014) Acceptance and attitudes toward a human-like socially assistive robot by older adults. Assist Technol 26(3):140–150

36. MacCallum RC, Widaman KF, Zhang S, Hong S (1999) Sample size in factor analysis. Psychol Methods 4(1):84–99

37. Majgaard G (2015) Humanoid robots in the classroom. IADIS Int J WWW/Internet 13(1):72–86

38. Marsh HW, Hau KT, Wen Z (2004) In search of golden rules: comment on hypothesis testing approaches to setting cut-off values for fit indexes and dangers in overgeneralizing findings. Struct Equ Model 11:320–341

39. Mubin O, Stevens C, Shahid S, Mahmud A (2013) A review of the applicability of robots in education. Technol Educ Learn 1:1–7

40. Nomura T, Kanda T, Suzuki T (2004) Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. Paper presented at the 3rd workshop on social intelligence design (SID2004), Twente

41. Nomura T, Kanda T, Suzuki T (2006) Experimental investigation into influence of negative attitudes toward robots on human–robot interaction. AI Soc 20(2):138–150

42. Nomura T, Kanda T, Suzuki T, Kato K (2006) Exploratory investigation into influence of negative attitudes toward robots on human–robot interaction. In: Lazinica A (ed) Mobile robots: towards new applications. I-Tech Education and Publishing, Rijeka, pp 215–232. https://doi.org/10.5772/4692

43. Nomura T, Kanda T, Suzuki T, Kato K (2008) Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots. IEEE Trans Rob 24(2):442–451

44. Nomura TT, Syrdal DS, Dautenhahn K (2015) Differences on social acceptance of humanoid robots between Japan and the UK. In: Proceedings of 4th int symposium on new frontiers in human–robot interaction. The Society for the Study of Artificial Intelligence and the Simulation of Behaviour (AISB)

45. Nomura T, Suzuki T, Kanda T, Kato K (2006) Measurement of negative attitudes toward robots. Interact Stud 7(3):437–454

46. Nomura T, Suzuki T, Kanda T, Kato K (2006b) Altered attitudes of people toward robots: investigation through the Negative Attitudes Toward Robots Scale. In: Proceedings of AAAI-06 workshop on human implications of human–robot interaction, vol 2006, pp 29–35

47. Nomura T, Shintani T, Fujii K, Hokabe K (2007) Experimental investigation of relationships between anxiety, negative attitudes, and allowable distance of robots. In: Proceedings of the 2nd IASTED international conference on human computer interaction, Chamonix, France. ACTA Press, pp 13–18

48. Prakash A, Rogers WA (2015) Why some humanoid faces are perceived more positively than others: effects of human-likeness and task. Int J Soc Robot 7(2):309–331

49. Ramachandran A, Scassellati B (2016) Long-term child–robot tutoring interactions: lessons learned. In: IEEE Ro-man workshop on long-term child–robot interaction, New York, August, vol 31

50. Raubenheimer J (2004) An item selection procedure to maximize scale reliability and validity. SA J Ind Psychol 30(4):59–64

51. Reich-Stiebert N, Eyssel F (2016) Robots in the classroom: what teachers think about teaching and learning with education robots. In: International conference on social robotics. Springer, Cham, pp 671–680

52. Ros R, Nalin M, Wood R, Baxter P, Looije R, Demiris Y et al (2011) Child–robot interaction in the wild: advice to the aspiring experimenter. In: Proceedings of the 13th international conference on multimodal interfaces. ACM, pp 335–342

53. Rosseel Y (2010) lavaan: an R package for structural equation modeling and more. Version 0.3-1. Retrieved, 29, 2010

54. Salter T, Michaud F, Larouche H (2010) How wild is wild? A taxonomy to characterize the 'wildness' of child-robot interaction. Int J Soc Robot 2:405–415

55. Serholt S, Barendregt W, Leite I, Hastie H, Jones A, Paiva A et al (2014) Teachers' views on the use of empathic robotic tutors in the classroom. In: The 23rd IEEE international symposium on robot and human interactive communication. IEEE, pp 955–960

56. Serholt S, Barendregt W, Vasalou A, Alves-Oliveira P, Jones A, Petisca S, Paiva A (2017) The case of classroom robots: teachers' deliberations on the ethical tensions. AI Soc 32(4):613–631

57. Syrdal DS, Dautenhahn K, Koay KL, Walters ML (2009) The negative attitudes towards robots scale and reactions to robot behaviour in a live human–robot interaction study. In: Proceedings of the AISB symposium on new frontiers in human–robot interaction. April 8–9. Edinburgh, UK, pp 109–115

58. Tay B, Jung Y, Park T (2014) When stereotypes meet robots: the double-edge sword of robot gender and personality in human–robot interaction. Comput Hum Behav 38:75–84

59. Teacher trends (National Center for Education Statistics). https://nces.ed.gov/fastfacts/display.asp?id=28

60. Tsui K, Desai M, Yanco H, Cramer H, Kemper N (2011) Measuring the perceptions of autonomous and known human controlled robots. Int J Intell Control Syst 16(2):1–16

61. Tung FW (2011) Influence of gender and age on the attitudes of children towards humanoid robots. In: International conference on human–computer interaction. Springer, Berlin, pp 637–646

62. van der Berghe R, Verhagen J, Oudgenoeg-Paz O, Ven SVD, Leseman P (2019) Social robots for language learning: a review. Rev Educ Res 89(2):259–295

63. Westlund JK, Gordon G, Spaulding S, Lee JJ, Plummer L, Martinez M et al (2016) Lessons from teachers on performing HRI studies with young children in schools. In: The eleventh ACM/IEEE international conference on human robot interaction. IEEE Press, pp 383–390

64. Worthington RL, Whittaker TA (2006) Scale development research: a content analysis and recommendations for best practices. Counsel Psychol 34(6):806–838

Yu Xia is a third-year doctoral student at Learning, Design, and Technology program at the Pennsylvania State University. Her current research interests lie in the fields computer-supported collaborative learning, technologically-enhanced learning, and human computer interaction.

Gerald LeTendre is the Harry Lawrence Batschelet II Chair of Educational Administration, and co-editor of The American Journal of Education. He received his B.A. (magna cum laude) in sociology from Harvard University and was awarded a Rockefeller Fellowship to do fieldwork among Tibetan refugees. He completed his graduate work

at Stanford University where he received his Master's (sociology) and Ph.D. (education). He has been the recipient of a Japan Foundation Fellowship, a Johann Jacobs Young Scholar Award, and a Spencer Post-doctoral Fellowship. He was a Fulbright Senior Scholar at the University of Bremen in Germany (2003–2004) and a Fulbright Research Scholar at Sophia University in Tokyo (2016). Dr. LeTendre has published on a broad range of topics in comparative and international education, His books include Learning to be Adolescent (Yale University Press), Improving Teacher Quality (Teachers College Press, co-authored with Motoko Akiba), Promoting and Sustaining a Quality Teacher Workforce (Emerald, co-edited with Alex Wiseman), and most recently The International Handbook of Teacher Quality and Policy (Routledge, co-edited with Motoko Akiba). His current research focuses on how information and communication technology (ICT)—specifically social robots—is changing teacher work roles and teacher professional development around the world.