Adapting Usability Metrics for a Socially Assistive, Kinesthetic, Mixed Reality Robot Tutoring Environment

Kartik Mahajan⋆, Thomas Groechel⋆, Roxanna Pakkar, Julia Cordero, Haemin Lee, and Maja J Matarić

Viterbi School of Engineering, University of Southern California, Los Angeles, California, USA {kmahajan, groechel, pakkar, jrcorder, haeminle, mataric}@usc.edu

Abstract. The field of Socially Assistive Robot (SAR) tutoring has extensively explored both subjective and objective usability metrics for seated tablet-based human-robot interactions. As SAR tutoring introduces kinesthetic mixed reality environments where students can move around and physically manipulate virtual objects, usability metrics for such interactions need to be re-evaluated. This paper applies standard usability metrics from seated 2D interactions to a kinesthetic mixed reality environment and validates those metrics with post-interaction survey data. Using data from a pilot study ($n = 9$) conducted with a mixed reality SAR tutor, three commonly-used metrics of usability for seated 2D tutoring interfaces were collected: **performance**, **manipulation time**, and **gaze**. The strength of each usability metric was compared to subjective survey-based scores measured with the System Usability Scale (SUS). The results show that usability scores were correlated with the gaze metric but not with the manipulation time or performance metrics. The findings provide interesting implications for the design and evaluation of kinesthetic mixed reality robot tutoring environments.

Keywords: socially assistive robot tutor · mixed reality · usability

1 Introduction

Socially Assistive Robot (SAR) tutoring has been extensively explored with a variety of users and usability studies [17]. Such robot tutors often rely on human-computer interfaces (e.g., tablets) to deliver content as well as increase the observability of the interaction without the need to rely on external sensors [3]. Consequently, usability studies in SAR tutoring typically focus on seated interaction and benefit from a reliable perceptual interface [3]. Advances in virtual, augmented, and mixed reality human-robot interaction (VAM-HRI) have enabled kinesthetic mixed reality environments where students move around and physically interact with coding blocks alongside a SAR tutor. Effective evaluation

⋆ equal contribution

of usability of these nascent 3D interfaces requires a re-evaluation of common usability metrics employed in other tutoring environments.

Usability has been studied in various tutoring environments using subjective and objective metrics. Subjective metrics include interview summaries [15] [19] and various survey tools, typically using Likert scales [12], such as the commonly used System Usability Scale (SUS), a 0-100 scale. Objective metrics include user performance, manipulation time, and gaze [6][1]. SUS is used for evaluating both on-line and in-person tutoring, while objective metrics are more commonly used in on-line tutoring. In SAR tutors, observability is typically limited; this challenge is even greater in kinesthetic environments, where students move around.

VAM-HRI represents an opportunity for obtaining objective behavioral metrics by providing a data-rich observable 3D interaction environment. Augmented reality head-mounted displays (ARHMDs), the common medium for VAM-HRI, do not rely on external sensors and can dynamically change the displayed environments [28]. This allows for synchronizing the robot tutor's and world states, simplifying on-line logging, and providing objective multimodal interaction data for analyzing the interaction.

This work applied objective usability metrics commonly used in seated 2D tutoring environments to data from a kinesthetic mixed reality environment pilot study ($n = 9$) and validated those metrics with post-interaction SUS survey data. Three different usability metrics were studied: 1) student performance via problem-solving policies; 2) object manipulation time; and 3) gaze concentration. The metrics were recorded over a 20 minute interaction involving a SAR tutor guiding a student through 7 coding exercises via a mixed reality visual programming language MoveToCode [7]. The strength of each usability metric was compared to subjective survey-based scores measured with the System Usability Scale (SUS). The results show that usability scores were correlated with the gaze metric but not with the manipulation time or performance metrics. The findings provide interesting implications for the design and evaluation of kinesthetic mixed reality robot tutoring environments.

2 Background and Related Work

2.1 Measuring Usability in Tutoring Systems

Usability, the ease of use and efficacy of a system [23], has been evaluated in various tutoring systems, ranging from on-line intelligent tutoring systems (ITS) to in-person SAR tutors. Across fields, usability is measured using qualitative and quantitative metrics of subjective post-interaction interviews [21][13] and self-report questionnaires [8] such as the System Usability Scale (SUS) [11].

In SAR tutor research, as well as in Web and interface-design, objective behavioral data are collected for usability analysis, such as eye gaze [16][2] and task completion time[27][24][20][5]. Objective behavioral metrics mitigate forms of reporting bias found in questionnaires [14]. As a validation metric, objective findings are often correlated with conclusions of post-interaction interviews, questionnaires, or study controls such as complexity of the interface [27][24][20][5].

Since a usable system has contributed to greater skill development [22], performance is a common usability metric. By nature, performance is evaluated differently based on tutoring environment. Clabaugh et al. measured math success by number of correct answers [4]. Roscoe et al. measured writing success based on a scaled essay score [22]. In the context of programming tutoring, because programming has been viewed as a multi-level problem-solving process, it has been evaluated individually at every stage, from exploration to submission [10]. Other techniques have evaluated programming solutions as a collection of policies, correcting each component, such as variable definition or if-statements [18].

This work explored applying objective usability metrics typically used in seated tutoring environments in a novel context of a kinesthetic mixed reality SAR tutor. The three metrics–manipulation time, and eye gaze–are compared with a standard subjective usability metric.

2.2 Measuring Usability in SAR and Mixed Reality Robot Tutoring

Virtual, augmented, and mixed reality for human-robot interaction (VAM-HRI) is a new and rapidly growing field of research [29]. Extending socially assistive robotics (SAR) tutors to VAM-HRI promises to significantly enhance interactivity as well as the collection of real-time user and usability data. SAR tutors often rely on human-computer interfaces (e.g., tablets) to deliver content as well as increase the observability of the interaction [3]. Objective behavioral data collection is often hindered by the lack of reliable yet unencumbering and unintrusive sensors [12]. Mixed reality tutors can enhance the learning experience by enabling a kinesthetic learning environment where students can move around and physically manipulate virtual objects [26] [9]. Currently, SAR tutoring systems typically focus on seated interactions and benefit from tablet interfaces [3]; in contrast, kinesthetic SAR tutoring is much more dynamic and calls for new usability metrics.

VAM provides a detailed, fully-controllable and observable interaction environments where reliable user behavioral data can be collected. Augmented reality head-mounted displays (ARHMDs), the common medium for VAM-HRI, readily synchronize with the robot's and world states, allowing for on-line logging [28]. Using augmented or mixed reality with a robot in the context of education is not new [30][25][31], but usability analysis in previous works is limited to subjective metrics [30]. Thus, this work explores applying usability metrics commonly used in seated 2D tutoring environemnts to the new context of kinesthetic mixed reality environments via VAM-HRI with SAR.

3 Dataset

The dataset used in this work was from a within subjects ($n = 9$) pilot study performed with a mixed reality visual programming language MoveToCode [7]

(Fig. 1) in which a SAR tutor aimed to increase a student's *kinesthetic curiosity* (KC), a metric involving the multimodal measure of a student's movement and curiosity. In the interaction, students combined *coding blocks* (e.g., if-blocks, print-blocks) by grabbing, dragging, and snapping blocks together in order to solve 7 beginner-level coding exercises. The acts of grabbing, dragging, and snapping blocks are part of the **manipulation time** metric, described in Sec. 4. Preset coding blocks were available to the participant at the beginning of each exercise. Tasks focused on building syntactic skills for integer addition, variable creation, and if-statements. The study and its results are under review for publication elsewhere.

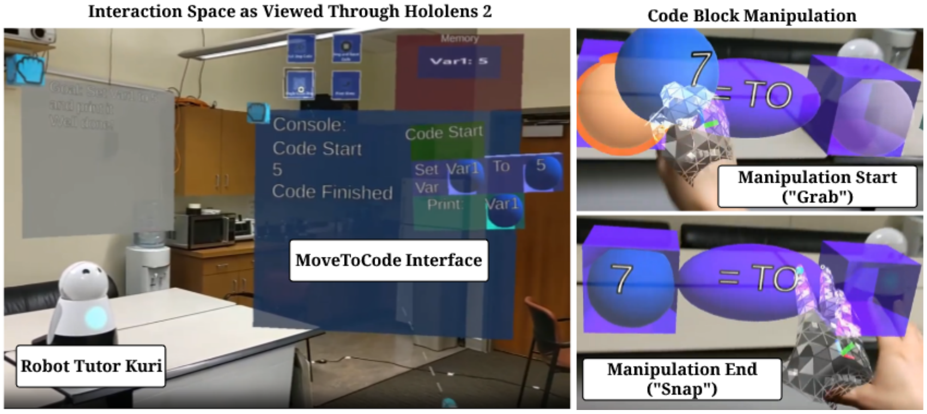

Fig. 1: MoveToCode (left) as seen by the participant through the Hololens 2. Code block manipulation (right) with a participant grabbing the block and letting it go to snap code blocks together.

The dataset includes 9 (2F,7M) of the 10 participants who were University of Southern California students with age range 19-27 ($\bar{x} = 22.8, \sigma = 2.9$). Participant 8 had two operating system crashes and was discarded from analysis. Behavioral data were collected at 0.02 sec intervals (50Hz) totaling 180 min of time series data yielding 540,000 rows. After the interaction, participants completed a ten-question questionnaire designed to measure individual SUS ratings.

This work examines the participants' objective and subjective metrics of usability. Specifically, it considers SUS survey results and the logged behavioral data of policy-evaluation, object manipulation time, and gaze concentration, described in the next section.

4 Usability Metrics

Since VAM-HRI for SAR tutoring is a nascent area, to understand usability metrics for kinesthetic mixed reality tutoring contexts, we reviewed usability studies of programming tutors [18] and web interfaces [27], and chose three commonly used reliable metrics: user performance, manipulation time, and gaze concentration.

Student Performance via Problem-Solving Policies: We measured a participant's performance by counting the number of good and bad policies created during a time frame. Introduced by Piech et al. [18], a *policy* is defined as any group of two or more code blocks. For example, if the participant was tasked with adding integer block 1 and integer block 2, a correct solution would include combining the two integer blocks with an addition block. A *Good* Policy (GP) for this task includes combining the integer block 1 with the addition block. A *Bad* Policy (BP) includes some other, incorrect step(s), such as combining the integer block 3 and the addition block.

Manipulation Time: MT is defined as the amount of time it takes a participant to grab a coding block and *snap* it to another block as can be seen in Fig. 1. *Snapping* was defined as the action grabbing a code block, dragging it to be in contact with another code block, and then releasing the currently held block to snap it to the contacted block. A successful manipulation event was logged from the time when an object was first grabbed (t_{grab}) to when an object was *snapped* to another object (t_{snap}).

Gaze Concentration: GC is defined as the amount of time a participant looked a 2D (x,y) pixel (i.e., cell) within the *interaction space* over a rolling time window tw_{GC}. The *interaction space*, shown in Fig. 1, included the MoveToCode interface and the physical robot tutor. The interaction space, measured in meters, was a 4m x 2.25m grid, totalling 3,600 cells measuring 0.05m x 0.05m each. A cell's score increased by 0.01 every frame the participant looked at that cell during tw_{GC}. The maximum cell score was capped at 1.

5 Results

5.1 Data Processing

All statistics were distributed between 0 and 1 using MinMax Scaling from Python's *sklearn* package (v0.24.2):

$$X_{scaled} = \sigma_x * (X_{max} - X_{min}) + X_{min} \tag{1}$$

where X_{scaled} is the new value for a data point in column X. To reduce skew of manipulation time (MT) results, a max MT of 10 seconds was empirically chosen, leaving 95.1% of the data. Any times over 10 seconds were adjusted to 10 seconds.

Not all participants completed all exercises; P5 failed to complete exercise 3 and P1,2,6,9 failed to complete exercise 6. This resulted in differently sized datasets for the different participants.

We calculated post-interaction SUS scores for all participants based on a 10-question survey, as shown in Fig. 2 ($\bar{x} = 53.06, \tilde{x} = 55.0, \sigma^x = 17.2, CV = 32.8\%$).

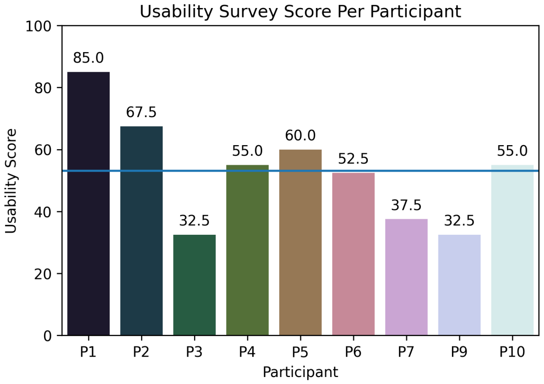

Fig. 2: SUS rating (0-100). The line indicates the median rating ($\tilde{x}$),

5.2 Unimodal Metric Analysis

We computed the variance and correlation of each metric to the SUS score (Fig. 2). A Levene's test was used to identify significant ($p < .05$) variance of each metric across participants, to signify the metric may distinguish different user behavior. We used Spearman's r correlation tests to validate the significance ($p < .05$) of each metric. We report the correlation of variance (CV) of each metric for unitless comparisons between metrics relative to their dispersion.

Policy-Evaluation Results To evaluate user performance, we recorded total Good Policies GP ($\bar{x} = 24.125, \sigma = 11.243, CV = 46.6\%$) and total Bad Policies BP ($\bar{x} = 3.625, \sigma = 3.24, CV = 89.2\%$) per participant and per exercise (Fig. 3). A Pearson's r test showed that there was no significant correlation between the total GP and BP ($r_p(9) = 0.581, p = .100$). A Levene's test indicated unequal variances per participant over exercises for total GP ($F = 6.72, p = .001$) yet no significant variance for total BP ($F = 0.993, p = .439$). This supports that GP may be effective in helping to differentiate user behavior, whereas BP may not be, due to its consistency across all participants. A Spearman's r correlation indicated no significant relationship between total GP and SUS score ($r_s(9) = -0.369, p = .327$), indicating total GP is not indicative of SUS score when observed unimodally. A Spearman's r correlation indicates no significant relationship between total BP and SUS score ($r_s(9) = -0.340, p = .370$), supporting that total BP is not indicative of SUS score.

These findings show that neither of the user performance metrics (GP or BP) alone were significant indicators of usability.

Manipulation Time Results To evaluate Manipulation Time MT, we recorded average MT per participant ($\bar{x} = 2.93s, \sigma^x = 0.76s, CV = 25.9\%$) and per exercise ($\bar{x} = 2.80s, \sigma^x = 0.57s, CV = 20.3\%$), as shown in Fig. 4. A

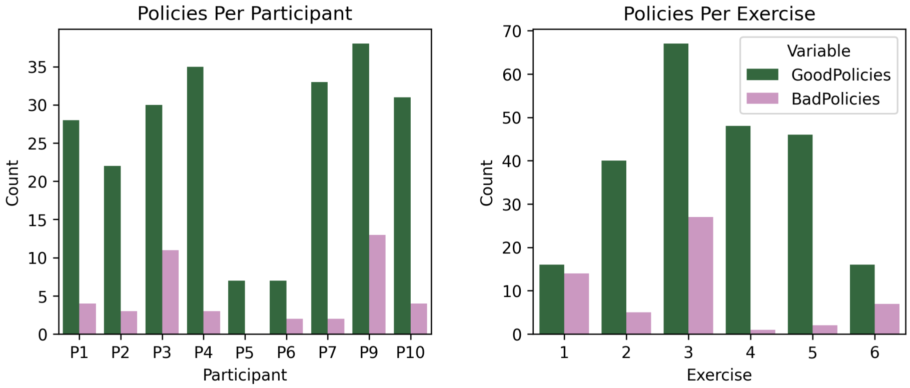

Fig. 3: Good Policies (GP) and Bad Policies(BP) viewed per participant and per interaction. Exercise 7 was free-play so there are no GP or BP recorded for it.

Levene's test indicated a significant variance among participant's average MT per exercise ($W = 5.94, p < .0001$), indicating MT is able to differentiate user behavior. A Spearman's r correlation indicated no significant correlation between average MT and SUS score ($r_s(9) = 0.353, p = .351$), indicating that average MT is not indicative of SUS score.

As an additional MT metric, we also calculated σ^{MT} ($\bar{x} = 3.51, \sigma^x = 0.816, CV = 23.2\%$). A Spearman's r correlation also showed no significant correlation between σ^{MT} and SUS score ($r_s(9) = -0.417, p = .263$), indicating that average σ^{MT} was not indicative of the SUS score.

These findings indicate that neither of the Manipulation Time (MT) metrics alone was a significant indicator of usability.

Eye Gaze Concentration Results To analyze eye gaze concentration GC, we empirically set tw_{GC} to 10 seconds. High intensity cell reads (HR) were defined as any cell with a value of 0.9 or higher, because 0.9 is over two standard deviations ($\sigma^{GC} = 0.306$) away from the average ($\bar{x} = 0.181$).

To evaluate HR, we recorded the total HR per participant ($\bar{x} = 153.44, \sigma^x = 23.733, CV = 15.4\%$) as shown in Fig. 5. A Levene's test showed a significant difference in variance among HR per time-step ($F = 38.1, p < .0001$), indicating that HR may distinguish user behavior. A Spearman's r correlation also indicates a significant relationship between total HR and the SUS score ($r_s(9) = 0.77, p = .014$), supporting that total HR is indicative of SUS score.

As an additional metric of GC, we calculated σ^{GC} ($\bar{x} = 23.42, \sigma = 14.22, CV = 15.4\%$), based on 2D coordinates of gaze, to represent how a participant's gaze traveled over a window. A Spearman's r correlation showed no significant relationship between σ^{GC} and the SUS score ($r_s(9) = 0.235, p = .542$), indicating that σ^{GC} is not indicative of SUS score.

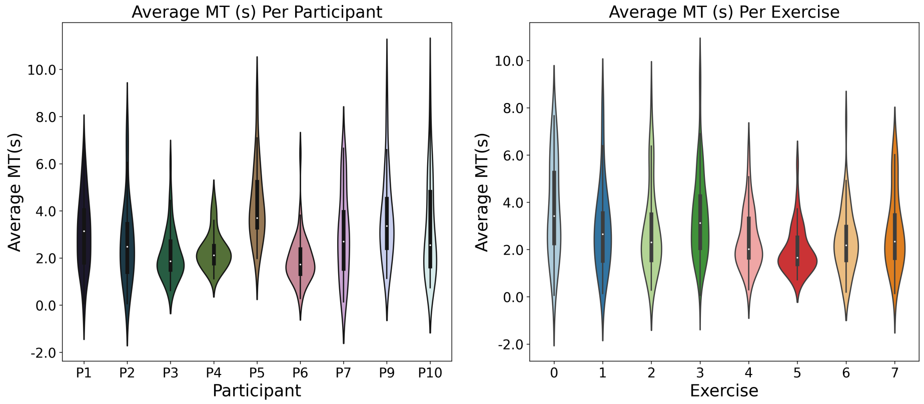

Fig. 4: Average Manipulation Time (MT) per exercise and per participant. MT records time between when a student chooses a coding block (e.g. if-statement, integer block) and snaps it to another component. Refer to Sec. 4 for more detail.

These findings indicate that HR was a significant metric for usability, whereas σ^{GC} was not.

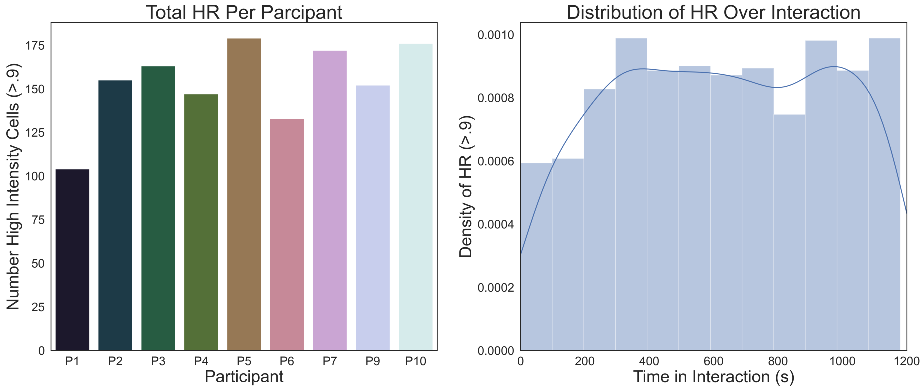

Fig. 5: Total High Intensity Cell Reads (score > 0.9) HR per participant recorded over a rolling time window $tw_{GC} = 10$. Cells are defined as any 2D pixel in the interaction space. For more details, see Sec. 4.

6 Discussion

The results showed varying degrees of success when applying common usability metrics in the context of kinesthetic mixed reality SAR tutoring. Per Spearman r tests, HR was the only metric that correlated with our SUS scores ($r_s(9) = 0.77, p = .014$) We hypothesize that this may be due to the varying sizes of our datasets ($Size_{GP/BP} = 289, Size_{MT} = 371, Size_{HR} = 12106$); GP, BP, and MT metrics yielded no significant results. The following insights can be drawn from this work.

Gaze concentration is a useful usability metric in VAM-HRI tutoring. As mentioned above, a Spearman r test revealed a significant ($p < .05$) correlation between HR and SUS scores ($r_s(9) = 0.77, p = .014$). A Pearson r test confirmed this correlation between HR and SUS, demonstrating the linearity of their correlation ($r_p(9) = 0.78, p = .011$).

Performance analysis may need to be re-evaluated as a metric in VAM-HRI tutoring. Policy-based evaluation yielded no correlation with SUS score($r_s(9) = -0.369, p = .327$). The accuracy of these findings is questionable given a significantly high CV (GP : 46.6%, BP : 89.2%) relative to that of HR (15.4%). Performance-based metrics have been successfully used in 2D tutoring environments [18] and should be further explored for VAM-HRI tutoring. Given the novelty of VAM-HRI for many users, we recommend using metrics of exploration, such as kinesthetic curiosity [7], that have demonstrated ability to distinguish participant behavior in VAM-HRI tutoring.

7 Conclusion

In summary, this work explored subjective and objective usability metrics typically used in seated 2D interactions in a kinesthetic mixed reality environment. Over a 20 minute pilot study ($n = 9$) conducted with a mixed reality SAR tutor, three commonly-used objective metrics of usability were collected–performance, manipulation time, and gaze–and where then correlated with a commonly used subjective metric System Usability Scale (SUS) metric. In the study, a mixed reality SAR tutor guided students through 7 beginner-level programming exercises via a mixed reality visual programming language MoveToCode [7] we developed. Subjective SUS scores were correlated with the objective gaze metric but not with the objective manipulation time or performance metrics. The findings serve to inform the design and evaluation of kinesthetic mixed reality robot tutoring environments.

Acknowledgements. This work was supported by NSF NRI 2.0 grant for "Communicate, Share, Adapt: A Mixed Reality Framework for Facilitating Robot Integration and Customization" (NSF IIS-1925083). We would also like to thank Matthew Rueben for all of his assistance.

References

1. Caleb-Solly, P., Dogramadzi, S., Huijnen, C.A., Heuvel, H.v.d.: Exploiting ability for human adaptation to facilitate improved human-robot interaction and acceptance. The Information Society **34**(3), 153–165 (2018)

2. Cho, H., Powell, D., Pichon, A., Kuhns, L.M., Garofalo, R., Schnall, R.: Eye-tracking retrospective think-aloud as a novel approach for a usability evaluation. International journal of medical informatics **129**, 366–373 (2019)
3. Clabaugh, C., Matarić, M.: Escaping oz: Autonomy in socially assistive robotics. Annual Review of Control, Robotics, and Autonomous Systems **2**, 33–61 (2019)
4. Clabaugh, C.E., Mahajan, K., Jain, S., Pakkar, R., Becerra, D., Shi, Z., Deng, E., Lee, R., Ragusa, G., Mataric, M.: Long-term personalization of an in-home socially assistive robot for children with autism spectrum disorders. Frontiers in Robotics and AI **6**, 110 (2019)
5. Dey, A., Billinghurst, M., Lindeman, R.W., Swan II, J.E.: A systematic review of usability studies in augmented reality between 2005 and 2014. In: 2016 IEEE international symposium on mixed and augmented reality (ISMAR-Adjunct). pp. 49–50. IEEE (2016)
6. Feingold-Polak, R., Elishay, A., Shahar, Y., Stein, M., Edan, Y., Levy-Tzedek, S.: Differences between young and old users when interacting with a humanoid robot: A qualitative usability study. Paladyn, Journal of Behavioral Robotics **9**(1), 183–192 (2018)
7. Groechel, T., Kuo, C., Dasgupta, R., Wathieu, A.: interaction-lab/movetocode: Doi release (Jun 2020). https://doi.org/10.5281/zenodo.3924514, `https://doi.org/10.5281/zenodo.3924514`
8. Holden, R.J., Campbell, N.L., Abebe, E., Clark, D.O., Ferguson, D., Bodke, K., Boustani, M.A., Callahan, C.M., et al.: Usability and feasibility of consumer-facing technology to reduce unsafe medication use by older adults. Research in Social and Administrative Pharmacy **16**(1), 54–61 (2020)
9. Ibrahim, R.H., Hussein, D.A.: Assessment of visual, auditory, and kinesthetic learning style among undergraduate nursing students. Int J Adv Nurs Stud **5**, 1–4 (2016)
10. Ichinco, M., Harms, K.J., Kelleher, C.: Towards understanding successful novice example user in blocks-based programming. Journal of Visual Languages and Sentient Systems **3**, 101–118 (2017)
11. Kaya, A., Ozturk, R., Gumussoy, C.A.: Usability measurement of mobile applications with system usability scale (sus). In: Industrial Engineering in the Big Data Era, pp. 389–400. Springer (2019)
12. Keizer, R.A.O., Van Velsen, L., Moncharmont, M., Riche, B., Ammour, N., Del Signore, S., Zia, G., Hermens, H., N'Dja, A.: Using socially assistive robots for monitoring and preventing frailty among older adults: a study on usability and user experience challenges. Health and Technology **9**(4), 595–605 (2019)
13. Lee, W.H., Lee, H.K.: The usability attributes and evaluation measurements of mobile media ar (augmented reality). Cogent Arts & Humanities **3**(1), 1241171 (2016)
14. Linek, S.B.: Order effects in usability questionnaires. Journal of Usability Studies **12**(4), 164–182 (2017)
15. Malik, N.A., Hanapiah, F.A., Rahman, R.A.A., Yussof, H.: Emergence of socially assistive robotics in rehabilitation for children with cerebral palsy: A review. International Journal of Advanced Robotic Systems **13**(3), 135 (2016)
16. Menges, R., Tamimi, H., Kumar, C., Walber, T., Schaefer, C., Staab, S.: Enhanced representation of web pages for usability analysis with eye tracking. In: Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications. pp. 1–9 (2018)

17. Papadopoulos, I., Lazzarino, R., Miah, S., Weaver, T., Thomas, B., Koulouglioti, C.: A systematic review of the literature regarding socially assistive robots in pre-tertiary education. Computers Education **155**, 103924 (2020). https://doi.org/https://doi.org/10.1016/j.compedu.2020.103924, `http://www.sciencedirect.com/science/article/pii/S0360131520301238`
18. Piech, C., Sahami, M., Huang, J., Guibas, L.: Autonomously generating hints by inferring problem solving policies. In: Proceedings of the second (2015) acm conference on learning@ scale. pp. 195–204 (2015)
19. Pino, M., Boulay, M., Jouen, F., Rigaud, A.S.: "are we ready for robots that care for us?" attitudes and opinions of older adults toward socially assistive robots. Frontiers in aging neuroscience **7**, 141 (2015)
20. Pranoto, H., Tho, C., Warnars, H.L.H.S., Abdurachman, E., Gaol, F.L., Soewito, B.: Usability testing method in augmented reality application. In: 2017 International Conference on Information Management and Technology (ICIMTech). pp. 181–186. IEEE (2017)
21. Rodriguez, R.G., Monteoliva, J.M., Pattini, A.E.: A comparative field usability study of two lighting measurement protocols. International Journal of Human Factors and Ergonomics **5**(4), 323–343 (2018)
22. Roscoe, R.D., Allen, L.K., Weston, J.L., Crossley, S.A., McNamara, D.S.: The writing pal intelligent tutoring system: Usability testing and development. Computers and Composition **34**, 39–59 (2014)
23. Shackel, B.: Usability–context, framework, definition, design and evaluation. Interacting with computers **21**(5-6), 339–346 (2009)
24. Sonderegger, A., Schmutz, S., Sauer, J.: The influence of age in usability testing. Applied Ergonomics **52**, 291–300 (2016)
25. Stein, G., Lédeczi, A.: Mixed reality robotics for stem education. In: 2019 IEEE Blocks and Beyond Workshop (B&B). pp. 49–53 (2019)
26. Vázquez, C., Xia, L., Aikawa, T., Maes, P.: Words in motion: Kinesthetic language learning in virtual reality. In: 2018 IEEE 18th International Conference on advanced learning technologies (ICALT). pp. 272–276. IEEE (2018)
27. Wang, J., Antonenko, P., Celepkolu, M., Jimenez, Y., Fieldman, E., Fieldman, A.: Exploring relationships between eye tracking and traditional usability testing data. International Journal of Human–Computer Interaction **35**(6), 483–494 (2019)
28. Williams, T., Hirshfield, L., Tran, N., Grant, T., Woodward, N.: Using augmented reality to better study human-robot interaction. In: HCII Conference on Virtual, Augmented, and Mixed Reality (2020)
29. Williams, T., Szafir, D., Chakraborti, T., Soh Khim, O., Rosen, E., Booth, S., Groechel, T.: Virtual, augmented, and mixed reality for human-robot interaction (vam-hri). In: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction. p. 663–664. HRI '20, Association for Computing Machinery, New York, NY, USA (2020). https://doi.org/10.1145/3371382.3374850, `https://doi.org/10.1145/3371382.3374850`
30. Xefteris, S., Palaigeorgiou, G.: Mixing educational robotics, tangibles and mixed reality environments for the interdisciplinary learning of geography and history (2019)
31. Yang, F.C.O.: The design of ar-based virtual educational robotics learning system. In: 2019 8th International Congress on Advanced Applied Informatics (IIAI-AAI). pp. 1055–1056. IEEE (2019)