Original research

Using actors to develop technologies for older adults with dementia: A pilot study

Jennifer Boger MASc PEng
Department of Occupational Science and Occupational Therapy,
University of Toronto & Toronto Rehabilitation Institute,
Toronto, Canada
E: jen.boger@utoronto.ca

Jesse Hoey, PhD
Department of Computing
University of Dundee, Scotland, United Kingdom

Kate Fenton MA
Department of Occupational Science and Occupational Therapy,
University of Toronto, Toronto, Canada

Tammy Craig BSc
Department of Occupational Science and Occupational Therapy,
University of Toronto, Toronto, Canada

Alex Mihailidis PhD PEng
Department of Occupational Science and Occupational Therapy &
Institute of Biomaterials and Biomedical Engineering, University of
Toronto, & Toronto Rehabilitation Institute, Toronto, Canada
E: alex.mihailidis@utoronto.ca

J. Boger, J. Hoey, K. Fenton, T. Craig, A. Mihailidis. Using actors to develop technologies for older adults with dementia: A pilot study. Gerontechnology 2010; 9(4):450-463 doi:10.4017/gt.2010.09.04.001.00 Emerging assistive technologies represent a potential resource for supporting older adults with dementia and their families. However, developing useful and effective technologies is challenging as it is particularly difficult to run prototype tests with potential users from this population; tests that are necessary to ensure the unique and complex needs of this targeted user group are appropriately complemented. This research investigates the possibility of using actors to simulate older adults with dementia to optimise technologies before they are used in clinical trials, thus potentially circumventing considerable time, financial, and logistical obstacles. To gauge the applicability of this approach from the perspectives of human experts and assistive technology, examples of six older adults with dementia and six actors simulating older adults with dementia participating in the task of handwashing were shown to professional caregivers and to COACH (a computer-based assistive technology). Data from believability rating tasks by the professional caregivers and interactions with COACH were compared for the older adult with dementia and actor groups. Results were promising with both the caregivers and COACH showing little difference between the people with dementia and the actors. While these preliminary findings are encouraging, the small sample size of this pilot study necessitates further research before definitive conclusions can be made.

Keywords: actors, older adults, dementia, assistive technology

Older adults are the fastest growing global demographic, with the number of people aged 60 years and older expected to almost triple from 606 million in 2000 to nearly 1.9 billion by 2050[1]. A large increase is predicted for people over 80 years of age from 69

million in 2000 to 377 million in 2050[1]. This will result in a change in the global demographic of people over the age of 60 from 19 percent in 2003 to 32 percent in 2050[1]. The likelihood and severity of age-related morbidities significantly increases as people age. Dementia, a disease that has profound implications for effected people and their families, currently affects one in ten people aged 65 or older and one in three people over the age of 85[2,3]. Coupled with a globally increasing average lifespan, dementia prevalence is predicted to double every 20 years, from 35.6 million people in 2010 to 115.4 million by 2050[4]. In North America alone, this growth is equivalent to a new dementia diagnosis every 7 seconds[5]. Dementia represents the leading cause of dependency and disability and is associated with significantly higher care costs than other chronic illnesses. For example, in the United Kingdom formal and informal care costs for dementia were estimated to be greater than those for stroke, heart disease and cancer combined[4].

Dementia compromises memory functions, including one's ability to remember when and how to complete tasks. This makes activities of daily living (ADL), such as washing and dressing, difficult or impossible to complete, thus greatly impacting an individual's ability to remain independent. Caring for someone with dementia not only requires time and effort on the part of the caregiver, but represents a fundamental change in the relationship between the caregiver (who is usually a spouse or an adult child) and the care recipient[6-8]. Not surprisingly, most people wish to age-in-place in their own homes, however, the emotional, physical, and financial burdens of dementia increase as the disease progresses resulting in increasingly difficult challenges for the caregiver, care recipient, and people in their social network. In many cases, there comes a point when the caregiver can no longer cope and the person with dementia is admitted to a long-term care facility, which is often devastating for both the caregiver and care recipient. In the United States, long-term care facilities for older adults with dementia are not only already filled to capacity, but are also expensive to run; in 2007 the estimated cost for nursing home care alone in the United States was 123 billion US$[9]. In 2001, less than 1% of people 65 to 69 years old were in long-term care compared to more than 43% of people 95 years or older. This reflects the fact that the majority of all older adults who require some form of care receive informal (non-paid, usually family-based) care within their community[10]. Clearly, delaying the placement of older adults with dementia into long-term care not only helps to solve social and logistical dilemmas, but would have a significant impact on mounting financial healthcare costs as a result of an aging population.

When creating any product or device, it is crucial that the design complements the needs and abilities of the targeted users. This is especially true for technologies aimed at supporting older adults with dementia and their caregivers, who generally have limited time and/or capabilities to learn how to use a new device. Technologies intended for use by older adults with dementia must support diminished cognitive abilities (for instance, little or no capacity for learning, poor memory for where objects are left, difficulties remembering how to use objects and tools, etc.), while being sensitive to other disabilities that often accompany aging, such as motor, vision, and auditory impairments. Therefore, to ensure they are accessible, technologies for this user group should be zero-effort (i.e., require no or little effort to operate). Zero-effort technology is as important to overburdened (and often elderly) caregivers as it is to people with dementia themselves.

User-centred design (UCD) is a method employed to capture the wants and needs of targeted device and product users[11]. In an effort to ensure the resulting product is appropriate, useful, and accepted, UCD includes people from the population(s) of interest at all stages in the device or product's development; from the initial planning stages through the

design and building stages to the testing of the prototype. However, the UCD approach requires a lengthy, iterative design process, which is often costly, disruptive to participants' schedules, and time consuming. This has been found to be especially true when involving users with disabilities[12-14]. Involving older adults with dementia in this process can be even more difficult because of increased financial requirements (for instance, relatively large number of research personnel required, long amounts of time required to complete the study, etc.), extensive ethical approval and consent procedures, as well as limitations on study locations (because, since the study participants cannot travel, the researchers must go to them)[15]. In addition to the significant burden placed on the people involved in prototype testing, the frail health of the participants means they can only participate in one trial per day or less, resulting in trials that could be completed in days by healthy individuals to be stretched out over many weeks, and sometimes months, with older adults with dementia participants. Additionally, health complications will often cause a participant to drop out of the (often multi-month long) study, resulting in lost data and/or the recruitment of another participant.

Several different types of assistive technologies have been developed in an effort to support aging-in-place for people with dementia. One such technology is the COACH system, which is designed to autonomously guide older adults with dementia through ADL[16]. COACH employs computer vision to track a user and objects of interest as s/he performs an ADL and provides audio and/or visual assistance if, and only if, the user requires it (for instance, s/he performs a step in the activity out of sequence, gets sidetracked, or is not sure what to do next). Artificial intelligence methods are employed to learn characteristics about each individual over time, such as what his/her level of independence is, what type of prompts are most effective for him/her, and the average amount of time it takes him/her to complete each step in the activity. Not only does this approach enable COACH

to customise guidance to each individual's needs, but also allows the system to adapt over time to changes in individuals' responsiveness and capabilities. Progressively complex iterations of the COACH system have been tested with older adults with dementia using the representative ADL of handwashing[16-18]. Handwashing was chosen because it is an activity one must do several times a day, it is a low-risk activity, and older adults with dementia often have trouble remembering what steps are involved and their proper sequence. Thus far, COACH development has adhered to UCD techniques, however, employing this method with older adults with dementia has caused development to be a lengthy and challenging endeavour.

Ideally, products and devices are optimised as much as possible before testing them with the population(s) of interest. Some of the challenges of involving population(s) of interest can be circumvented by optimising a prototype in simulated trials before testing it with the intended users. While it has been identified that disability simulation tests could be an easy, inexpensive, and potentially useful approach to identifying product deficiencies, published studies on this topic are quite sparse[19]. Using public information kiosks as a representative product, Law and Vanderheiden used 15 blind and 15 blindfolded people to investigate whether people simulating a disability would identify the same usability problems as people who had the disability in question[20]. The results are encouraging, with the number and types of problems discovered by each group being within 3% of each other. While simulating a disability may be a reasonable approach for many user groups, it poses a challenge for devices intended for people with dementia as these users exhibit behaviours that are quite difficult to replicate. Reliably and accurately replicating human behaviours is a task that is beyond the capabilities of most researchers, however, it is exactly what professional actors have trained extensively to do. Actors are widely used as simulated or standardised patients in the medical profes-

sion as a method of teaching and assessment[21,22]. Role-playing is an extensively used patient simulation technique where actors take on the part of patients and are encouraged to explore attitudes, feelings, and behaviours of their role to help medical students recognise and interact with the target patient group[23].

If actors could use role-playing to accurately reproduce the behaviours seen in older adults with dementia, then simulated trials involving actors might not only be used to improve the optimisation process of devices for people with dementia, but could also provide designers with the opportunity to investigate specific scenarios. This, in turn, could substantially reduce development time and financial costs while simultaneously improving device applicability and usability. However, in order for this approach to be successful the actors must be able to believably simulate older adults with dementia from the perspectives of both human experts and the devices they are interacting with.

OBJECTIVE
The objective of this study was to investigate the following research questions: (i) Can actors believably simulate older adults with dementia, and (ii) Can actors simulating older adults with dementia be used to optimise technologies before they go to clinical trials?

METHOD
The research presented in this paper investigated the applicability of using actors simulating older adults with dementia (hereafter referred to as *actors*) to develop technologies for older adults with dementia (hereafter referred to as *OAwD*) by having actors employ role-playing to simulate OAwD performing the ADL of handwashing. The believability of the actors from the standpoints of human experts and technology was subsequently investigated through two parts. In Part 1, professional caregivers watched several video clips of OAwD and actors participating in handwashing. Opinions regarding believability were captured

by having the caregivers rate the believability of the person in the video and by asking the caregivers to identify if the person was an actor or an OAwD. In Part 2, OAwD and actors interacted with COACH (described in the Introduction), which was used to investigate the effectiveness of actors to simulate OAwD within the context of a specific assistive technology. In this part of the study, responses from COACH to the actors were compared to those for OAwD, which were captured in a previous study[16].

Handwashing was selected as an illustrative activity because it is complex enough to present a challenge to most people with moderate-to-severe dementia, is a safe task to replicate (for instance, little or no risk of the participants hurting themselves or becoming unnecessarily upset), and there was previously-obtained handwashing data (both for caregiver- and device-guided scenarios) available for OAwD. Prior to commencing this research, approval for this study was received from the appropriate institutional research ethics boards. All statistical analysis were completed using SAS v9.1 for Windows (released in 2003 by SAS Institute Inc., Cary, NC, USA).

Dementia video footage
Videos of OAwD were taken from footage captured during a previous efficacy study with COACH that involved five female and one male participants[16]. In this previous study, trials were conducted using an A-B-A-B single-subject research design, with A-phases involving handwashing guidance by a human caregiver and B-phases involving handwashing guidance by COACH. Participants conducted 40 trials in total: 20 A-phase (human guided) and 20 B-phase (COACH guided) trials. Mini-Mental State Examination (MMSE) scores were used to gauge the participants' level of dementia[24]. Informed consent was obtained from the participants' substitute decision makers prior to using any of the OAwD's data. For more details about the previous COACH efficacy study, the reader is referred to[16,25].

Actor video footage

Recruitment, training, and direction of the actors was coordinated by a member of the research team who has a Masters degree in (Dramatic) Fine Arts as well as extensive experience with clinical trials involving OAwD. Actors were recruited from a group of professional actors over the age of 65. Interested actors were invited to audition, which included an interview of his/her acting experience, prior experience with older adults with dementia, and a short reading selected by the actor. After auditioning, seven actors were selected and were given a character outline that was based on one of the six OAwD from the previously-completed efficacy study. As there were seven actors studying six OAwD, two actors studied the same OAwD. After the actor studied his/her outline, the study coordinator coached each actor as s/he watched an hour of video footage of his/her OAwD counterpart and subsequently practiced simulating handwashing for an hour. Within a week of his/her coaching session, the actor simulated the OAwD s/he studied in 45 handwashing trials: five practice trials, 20 trials guided by the same human caregiver who guided the OAwD in the previous study, and 20 trials guided by COACH. Actors interacted with the same version of COACH that was used in the previously completed COACH efficacy study (i.e., the same version of COACH the OAwD used). Actors had no exposure to COACH prior to the simulation trials (i.e., the actor had never seen COACH or videos of OAwD interacting with COACH). The trials employed the washroom and equipment setup that was used in[16]. It was decided *a priori* that one of the actors' footage would be discarded in order to balance the groups (i.e., five female and one male actors to balance the already-obtained five female and one male OAwD footage), thus a female actor's footage was discarded based on slightly overly-quick mannerisms and vocal lucidity. All actors were compensated for their time.

Evaluation by professional caregivers

The first research question attempted to determine if actors could simulate handwashing in a manner that was comparable to OAwD. To accomplish this, professional caregivers who had clinical experience with older adults with dementia were asked to share their opinions regarding the believability of video clips of the actors and OAwD through two rating tasks, which are described below. Only footage of caregiver-guided (A-phase) trials were shown as it was felt COACH's novel guidance technique would be unnecessarily distracting. No video segment was seen by the same caregiver participant more than once. Informed consent and demographics were collected from each participant before s/he started the rating tasks. Part 1 took approximately an hour to complete.

Believability

The first task employed a method similar to the one used by Rosen et al., which examined the believability of actors simulating different stages of depression[26]. Ten of the 20 caregiver-guided trials were randomly selected for each of the OAwD/actors and edited into one-minute segments. As the number of segments depended on how quickly the OAwD/actor washed his/her hands, there were more segments for some OAwD/actors than others (on average there were 33 segments per OAwD/actor; min=17, max=57).

The caregiver participants were asked to view and rate 20 video segments; 10 OAwD segments and 10 actor segments. Video segments were selected for each caregiver participant by randomly choosing three OAwD and three actors and then randomly choosing three or four segments for each of these OAwD/actors. The resulting 20 video segments were then randomly presented to the caregiver participant. After viewing each segment, the caregiver participant was asked to rate the question "Was the client behaving in the same way that a person with dementia would?" using a Likert-scale with one corresponding to "not at all" and 10 corresponding to "very much". Space was given after each clip so that the caregiver participant could elaborate on why s/he

gave the score s/he did or provide any other comments s/he wished to make.

The above method resulted in a partially balanced, incomplete block design, with caregiver participants as blocks and OAwD/actors as treatments. The blocks are considered to be incomplete as each caregiver participant viewed clips from three of the six OAwD and three of the six actors. The remaining OAwD/actors were shown to the caregiver participant in Task 2 (described below) to prevent any opinions formed about the people seen by the caregiver participants in Task 1 biasing ratings in Task 2. Additionally, while each caregiver participant eventually saw segments of all the OAwD and actors, the random selection of which OAwD/actors were seen in Task 1 resulted in some OAwD/actors being rated more often in Task 1 than Task 2 and vice versa.

In the context of this study, intraclass correlation can be used to estimate whether or not different caregiver participants gave a similar score to the same OAwD/actor. The intraclass correlations were calculated for the dataset as a whole, as well as separately for the OAwD and actor groups (*Equation 1*).

$$\hat{R} = \hat{\sigma}_S^2 / (\hat{\sigma}_C^2 + \hat{\sigma}_S^2 + \hat{\sigma}_e^2) \qquad [1]$$

where C=Caregiver rater, S=OAwD/actor, and e=error.

Identification of the actor
The second task involved a direct comparison of the actor and OAwD conditions. For this task each caregiver participant was shown ten pairs of handwashing segments (for a total of 20 segments) selected from the three actors and three OAwD that the caregiver participant had not seen in Task 1. Each pair consisted of one OAwD segment and one actor segment of the same step in the handwashing task (i.e., turning on the water, getting soap, rinsing hands, drying hands, or turning off the water). OAwD were randomly paired with actors for each segment pair (i.e., the actor may or may

not have been paired with the OAwD s/he trained on). Each of the five handwashing steps was seen twice and presented to the participant in a random order. In five of the pairs the OAwD was shown first and in the other five the actor was viewed first, although the ordering of the pairs was random (i.e., the number of times in a row the caregiver participant viewed an actor first and vice versa was random). After viewing each pair, the caregiver participant was asked to identify which clip contained the actor and to provide any comments s/he wished to make. Task 2 is considered to be a partially balanced incomplete block design for the same reasons as Task 1.

Self-reported ratings
Upon the completion of the interview, the caregiver participant was asked to circle his/her answer to the question "When I was making my choices, I felt: very unsure, somewhat unsure, somewhat sure, or very sure". Space was given for the rater to share any general comments s/he had.

Interaction with COACH
The second research question investigated whether actors simulating dementia interact with assistive technology in the same manner that real people with dementia would in order to validate actor simulation as a feasible approach for device optimisation. Videos of the actor-simulated trials with COACH were reviewed and scored by the same research assistant using the same methods as the previously completed efficacy study[16]. Data from the actor-simulated trials were analysed and compared to those from the previously completed efficacy study.

Table 1. The four possible conditions that were used to evaluate COACH performance; for each action taken by the user, the corresponding COACH response was scored

COACH response	User action	
	Error	**No error**
Prompt	Hit	False alarm
No prompt	Miss	Correct reject

Device performance was analysed using the same methods and equations that were used to describe the performance of previous versions of COACH[16,17]. Hits, misses, false alarms, and correct rejects made by COACH were identified (*Table 1*). Using signal detection theory, these data were also used to calculate device sensitivity, a measure of COACH detecting an error when one was made, thus giving a prompt (*Equation 2*), specificity, a measure of COACH correctly identifying when the user is handwashing correctly, and thus not giving a prompt (*Equation 3*), and accuracy, a measure of how often COACH correctly interpreted the state of the device's environment (*Equation 4*).

$$Sensitivity = \frac{Hits}{Hits + Misses} \qquad [2]$$

$$Specificity = \frac{Correct\ Rejects}{False\ Alarms + Correct\ Rejects} \qquad [3]$$

$$Accuracy = \frac{Hits + Correct\ Rejects}{Misses + False\ Alarms} \qquad [4]$$

RESULTS
Participants
Eleven professional caregivers participated in this study. All participants were female, had an average age of 39.4 years (SD=10.3), and had an average of 6.5 years (SD=6.5) experience caring for older adults with dementia. All of the participants worked at a different institution from that of the OAwD and were therefore unfamiliar with all the people shown in the videos.

Believability
Rating scores given by the caregiver participants for Part 1, Task 1 were plotted against a normal density curve and agreement was considered to be close enough for the assumption of normally distributed data to hold. A statistically significant interaction effect was detected between the caregiver participants and the OAwD/actor status, i.e., whether the person in the video was an OAwD or an actor (*Table 2*). The least square means were used to model the data and were adjusted to take into consideration subject effects (*Table 3*).

Intraclass correlation values were calculated using Equation 1 to be 0.34 for the group as a whole, 0.35 for the OAwD group, and 0.43 for the actor group.

The overall average score given to each OAwD/actor was calculated (Figure 1). Note that while Figure 1 highlights the actor and the OAwD that s/he studied and replicated, the selection of which OAwD/actors were shown to the caregiver participant in Task 1 was random. Therefore, in Task 1 a caregiver participant may or may not have seen both the actor and the OAwD who s/he studied.

Actor identification
Caregiver participants' choices were scored as binary ratings reflecting whether or not the actor was identified. For the 109 responses recorded, 71 times (65.1%) the actor was correctly identified and 38 times (34.9%)

Table 2. Results of ANOVA analysis; df=Degrees of freedom; OAwD=Older adult with dementia; Actor= Adult actor simulating people with dementia; *=individual OAwD (5 df) and Actor (5 df)

Variation source	df	Σ squares	Mean square	F-ratio	p
Status OAwD / Actor	1	1.89	1.89	0.03	0.874
Caregiver rater	10	300.89	30.09	3.07	0.046
Rater x Status	10	98.09	9.81	2.38	0.011
Subject*	10	598.25	59.83	14.51	<0.001
Previously seen OAwD	1	31.58	31.58	7.66	0.006
Error	186	767.05	4.12		
Corrected total	218	1931.01			

Table 3. Raw and adjusted (for subject and interaction effects) least square means of the believability rating scores for professional caregivers (Rater) of video clips containing Older Adult with Dementia (OAwD) and adult actor simulating people with dementia (Actor); SE=Standard Error

| Rater	Mean			
	Raw (SE)		Adjusted least square (SE)	
	OAwD	Actor	OAwD	Actor
1	7.0 (1.12)	6.5 (1.20)	6.4 (0.69)	6.4 (0.69)
2	7.7 (0.83)	6.4 (0.96)	6.0 (0.69)	5.6 (0.68)
3	4.0 (0.65)	4.0 (0.60)	4.2 (0.69)	4.6 (0.70)
4	3.6 (0.73)	5.2 (0.85)	4.2 (0.69)	4.9 (0.69)
5	5.1 (0.82)	2.5 (0.50)	6.8 (0.69)	3.2 (0.69)
6	3.5 (0.86)	5.0 (0.82)	4.3 (0.70)	5.9 (0.69)
7	8.4 (0.79)	8.5 (0.37)	8.7 (0.68)	9.3 (0.70)
8	6.9 (1.10)	3.9 (0.95)	5.9 (0.68)	4.1 (0.72)
9	6.7 (0.68)	5.7 (0.47)	6.1 (0.68)	5.6 (0.69)
10	5.2 (1.32)	5.7 (0.83)	4.5 (0.68)	5.4 (0.69)
11	5.6 (0.67)	5.7 (0.73)	4.8 (0.69)	4.8 (0.69)
All raters	5.8 (0.30)	5.4 (0.27)	5.6 (0.20)	5.4 (0.21)
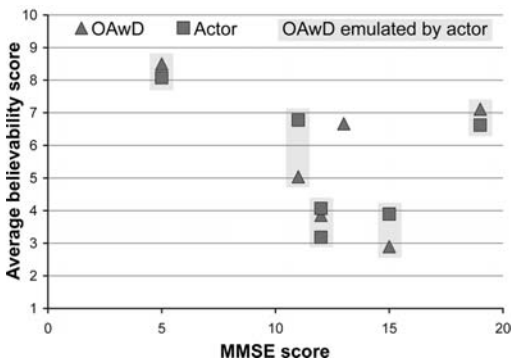
Figure 1. Average believability score received by each older adult with dementia (OAwD) and simulating actor connected by shaded areas; OAwD with MMSE (Mini-Mental State Examination) of 12 was studied by 2 actors; as established a priori to ensure a balanced sample size the actor who studied the OAwD with an MMSE of 13 was discarded as a result of overly-quick mannerisms and lucidity

the OAwD was mistaken for the actor. The data were arranged to examine if scoring was influenced by which subject the caregiver participant saw first in any given pair, i.e., whether the clip containing the OAwD was presented before or after the clip containing the actor (*Table 4*). While the average of correctly identified clips may be different (75.9% correct when OAwD was seen first

and 54.5% correct when the actor was seen first), it cannot be known conclusively if there is a significant difference as the data are too sparse to examine any possible interactions. Data were then analysed to see if responses differed from chance (i.e., 50/50). As the data were too sparse to implement two-way interaction terms and the number of distinct categories in the logistic-regression model were too great (107 categories with only 109 observations), stepwise regression was selected to investigate significant variables. As discussed above, there may be an interaction between which clip was seen first, therefore the 50/50 hypothesis was tested separately on the OAwD and actor groups as well as on the data set as a whole (*Equation 5*). The higher z-values represent a greater likelihood that the caregiver participants' scores differed from 50%, i.e., chance (*Table 5*).

Testing: $H_0: p = 0.5$ and $H_a: p \neq 0.5$ [5]

Using: $z_{obs} = \dfrac{\hat{p} - 0.50}{\sqrt{\dfrac{\hat{p}(1-\hat{p})}{n}}}$

Table 4. Percentage of correct responses from the professional caregiver (Rater) when asked to identify which clip contained an actor; OAwD=Older adults with dementia; Actor= Adult actor simulating people with dementia; *= calculated using the raw dataset with one observation missing (n=109 instead of n=110)

Rater	% correct		Overall
	Clip viewed first		
	OAwD	Actor	
1	66.7	57.1	60
2	100.0	50.0	80
3	40.0	80.0	60
4	100.0	40.0	70
5	80.0	40.0	60
6	40.0	20.0	30
7	80.0	60.0	70
8	80.0	40.0	60
9	60.0	60.0	60
10	100.0	60.0	80
11	80.0	100.0	89
Overall*	75.9	54.5	65

Where:
$\hat{p}$=percentage correct
n=number of observations

Self-reported caregiver participants' feelings of certainty were compiled (*Figure 2*). None of the caregivers felt very sure, most of them were somewhat unsure.
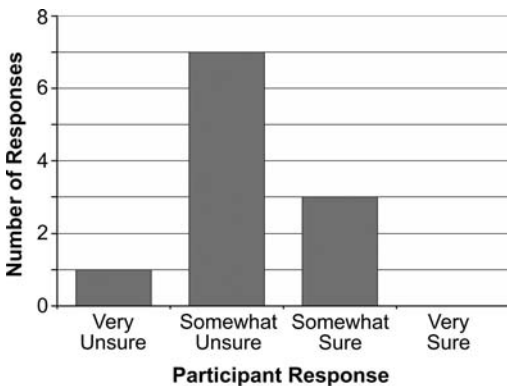
Table 5. Caregiver scores tested against chance (50/50); OAwD=Older adults with dementia; Actor= Adult actor simulating people with dementia; CI=Confidence interval

Characteristic	First shown		All pairs
	OAwD	Actor	
n	54	55	109
% correct	75.9	54.5	65.1
z-value	4.46	0.68	3.32
p	<0.001	0.249	<0.001
95% CI, %	64.5-87.3	41.4-67.7	56.2-74.1

Table 6. Characteristics for older adults with dementia (OAwD) and actor groups simulating dementia (Actor) when interacting with COACH , as calculated using Equations 2, 3 and 4

Characteristic	OAwD	Actor
Sensitivity	0.89	0.84
Specificity	0.74	0.64
Accuracy	3.41	2.27

Interactions with COACH

For each step in the handwashing task (for instance, getting the soap, drying hands, etc.), actions by COACH and the reaction of the OAwD/actor to these actions was recorded (*Figure 3*). In 120 trials, COACH took 750 actions for the OAwD group and 864 actions for the actor group. Sensitivity, specificity, and accuracy for both the actors and OAwD were calculated using Equations 2, 3, and 4 (*Table 6*).
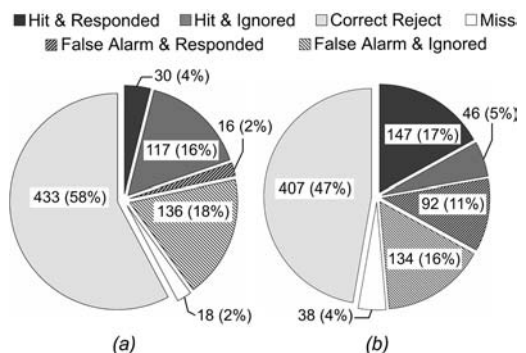

Figure 2. Caregiver' self-reported feelings of certainty regarding scores given during rating tasks

Figure 3. Number of COACH actions and subsequent reactions for (a) Older adults with dementia, and (b) Actors simulating their dementia

DISCUSSION

The ANOVA calculated for Part 1, Task 1 showed a statistically significant interaction effect between the rater and OAwD/actor status. However, looking at the least square mean (*Table 3*), it can be seen that after the data model was adjusted for subject effects eight of the raters gave similar scores to both the OAwD and actor groups, while the remaining three raters were split between consistently higher (C5 and C8) or lower (C6) scores. Perhaps C5 and C8 were particularly astute and C6 was less so, although it seems equally reasonable that scoring deviations were a by-product of the specific (and different) clips the raters viewed, and therefore, due to chance. With this explanation for variability in mind, since the majority of the caregiver participants assigned similar scores, the overall mean for the OAwD group could be considered to be the same as for the actor group.

The hypothesis that the average believability scores of the OAwD and actor groups are comparable is reinforced by the data in (*Figure 1*). Interestingly, while believability ratings do not appear to be related to the OAwD's level of dementia (as represented by the MMSE score), an actor's believability score does seem to mimic the one given to the OAwD that s/he studied. Additionally, there appears to be no trend as to whether the OAwD or the actor who studied him/her received a higher believability score. This may indicate that actors can not only simulate OAwD, but that they can believably replicate specific behaviours and mannerisms. Also, since believability scores do not seem to reflect MMSE values, this suggests that a person's believability score likely reflects some combination of functional abilities and mannerisms during handwashing rather than dementia level alone. This is not a surprising outcome as individuals' ADL competency and approach to executing ADL steps varies greatly. As such, while there is an overall trend toward poorer ADL completion with higher levels of dementia, the two are not tightly correlated. Comments made

by the caregiver participants during this task support this line of thought, as comments generally focused on mannerisms rather than the person's abilities or capabilities. Examples of comments include: "[I] don't believe residents act that way", about OAwD with an MMSE of 15; "Didn't look real", about an actor simulating an OAwD with an MMSE of 15; "Overdoing it! Trying too hard", about OAwD with an MMSE of 19; and "Just looked very real", about actor simulating an OAwD with an MMSE of 12.

Although slightly higher for the actor group, the intraclass correlation appears to be similar when compared for the three conditions that were examined (the dataset as a whole, the OAwD group alone, and the actor group alone). From this, one can surmise that the intraclass correlation, and therefore the variability of the ratings given by caregiver participants for the same OAwD/actor, is not significantly different between the OAwD and actor groups or within the dataset as a whole. While the intraclass correlations for all three conditions are similar, they are also rather low, signifying that while the amount of variability within an OAwD/actor is similar to the variability within the other OAwD/actors, the amount of variability within each OAwD/actor is significant (i.e., in general, there is little or no "clustering" of the ratings given to each OAwD/actor). This is not an unexpected result as people with dementia routinely exhibit high levels of variation in their behaviour, mannerisms, and functional capabilities on a daily, and sometimes hourly, basis as a result of a myriad of compounding factors, such as medications, illness, fatigue, and the nature of dementia itself. One might speculate that the slightly higher intraclass correlation value for the actor group signifies that actors demonstrate slightly more repeatable behaviours than OAwDs, however, the data are too sparse to make any significant conclusions.

The analysis conducted on the data from Part 1, Task 2 shows that there may be an effect depending on whether the OAwD or the actor clip is seen first in the pair. Although the

data were too sparse to conduct a compelling statistical analysis regarding this effect, the stepwise regression analysis (*Table 5*) shows that the caregiver participant identified the actor correctly more than half the time when the OAwD was seen first, but did not when the actor was seen first. This translates into the caregiver participants correctly identifying the actor 65% of the time when the group was analysed as a whole. This suggests that professional caregivers were at least somewhat able to pick out which clip contained the actor, although their ability to do so seemed to be impacted by whether they saw the OAwD or the actor first. This tendency may be explained by the phenomenon of belief updating. As outlined by Hogarth and Einhorn[27], a 'primacy' effect can be established by the first 'option' that is presented. This bias can be present in medical diagnostics[28]. In this study, the caregiver participant was asked to identify the actor by checking a box next to the appropriate answer to the statement: "The actor was in: The first video or The second video". By asking the participant to focus on the fact that there was an actor involved in one of the videos may have primed the participants. Moreover, by asking the raters to identify an actor may have resulted in the participants examining each video as though it contained an actor, causing any discrepancies to become more prevalent when the actor was in the second video in the pair. To remove this bias, any future studies that wish to test if the actor can be identified could ask half the participants to identify which videos contain the actors, and the other half of the participants which videos contain the OAwD. Regardless of the cause, while the overall 65% positive identification value was shown to be statistically significantly different from 50% (which would be equal to chance), it is still distant from 100%.

Most of the caregiver participants were 'somewhat unsure' of their answers in their self-reported certainty ratings. Comments made by the caregiver participants during Part 1 supported this outcome. For instance, one caregiver participant began searching for clues, such as nail polish, identification tags, or hair being messy (all of which were replicated with the actors, and therefore not unique to either group). Also, the large majority of the comments made by the caregiver participants focused on strategies employed by the caregiver and/or the actions of the care recipients in the videos, as opposed to OAwD/actor believability or scepticism thereof.

The results of the statistical analysis performed on the data for Part 1, Task 1 suggest that the mean and variability of the OAwD and actor groups are comparable, therefore it is reasonable to suppose that the raters felt the clips they saw to be equally believable. When this result is coupled with the results from Tasks 2 and 3, the overall data from Part 1 tend to support the hypothesis that older adult actors can simulate older adults with dementia believably enough that professional caregivers rate the two groups similarly, and perhaps even equally. While this outcome is encouraging, it must be kept in mind that the data reported above are too sparse to come to any strong conclusions.

For Part 2, COACH demonstrated quite a similar response for 'hits' (20% for OAwD, 22% for actors) and 'misses' (2% for OAwD, 4% for actors). While the 'false alarm' (20% for OAwD, 27% for actors) and 'correct' reject (58% for OAwD, 47% for actors) rates showed a greater separation, at less than 10% difference, this is considered to be a tolerable result particularly when considering the small sample size. Moreover, the actors tended to be more responsive to prompts given by COACH, as can be seen in the 'hit & responded' (4% for OAwD, 17% for actors) and 'false alarm & responded' (2% for OAwD, 11% for actors) rates. It is interesting that the overall false alarm was lower and correct reject rate was higher for OAwD than for actors. This is reflected in correspondingly higher sensitivity, specificity, and accuracy values, which suggests that the COACH system works better with OAwD than it does with actors. While this

could well be because COACH was optimised through trials with OAwD, it also translates into some important considerations regarding using actors for device development. Should designers want to gain a true perspective, then actors must elicit the same responses from the devices they interact with as the target population would. In general, it was observed that the actors responded more often, quickly, and purposefully to prompts from COACH than OAwD did. In the future, this over-reaction might be mitigated through more thorough training of the actors and possibly exposure to the device they will be interacting with before they are asked to emulate dementia. It is possible that a device that responds well to actors will respond just as well, if not better, to OAwD, although the opposite scenario is equally realistic. While the actors themselves may have been more responsive to prompts from COACH, overall the actions taken by COACH appear to follow the same general pattern for both actors and OAwD.

While this study supports the concept that actors can be used to optimise assistive technologies for older adults with dementia before testing it with the target population, it must be kept in mind that this pilot work only examines one assistive technology in a highly-controlled setting. Studies involving more actors and older adults with dementia, and thus more data collection and analysis, must be done to gain conclusive evidence one way or another. In addition to a larger sample size, other tasks and activities must be emulated. Other discrepancies, such as functional abilities and comorbidities, should be measured in addition to MMSE scores to enable a more in-depth comparison. Future studies also need to examine different types of assistive technologies. Other factors, such as the amount of training given to the actors and whether or not they should be familiarised with the technology(ies) prior to interaction, are areas that need to be investigated. For at least a portion of the work, care should be taken to ensure the raters view a balanced sample of videos so that more comparable results may be drawn. The most conclusive work could be to optimise an assistive technology under two conditions, one version using older adults with dementia and the other using actors, and compare the performance of the two versions when they are used by older adults who have dementia. Finally, developers who choose to employ simulated trials must identify at what stage in the device development process the convenience of using actors is surpassed by the need to involve trials with the true population of interest.

Conclusions

While a final prototype must always be tested with the population in question, this pilot work suggests that using professional older adult actors to optimise assistive technologies for older adults with dementia may constitute a viable alternative to clinical trials alone. Indeed, the results elicited from COACH by the actors are encouraging considering that the version of COACH used in this research was the product of three previous iterations of clinical testing with older adults with dementia, which took place over the course of several years. Should it prove to be applicable, the approach investigated through this research could result in significant savings in terms of development time, costs, and demands on the clinical population. Moreover, using actors may enable significantly greater access to dementia-behaviours, overcoming many of the current ethical, time, and resource constraints that surround trials with a clinical population. Ultimately, this could translate into the more efficient fabrication of a greater number of appropriate and useful technologies, thus getting new and effective supportive devices to the people that need them more quickly. While this research examined a high-tech assistive device, it is plausible that actors could be used in the development of low-tech devices and other related care areas, such as the formulation of care strategies and training of clinicians.

References

1. United Nations. World Population Prospects: The 2002 Revision. New York: United Nations 2003
2. Ablitt A, Jones GV, Muers J. Living with dementia: A systematic review of the influence of relationship factors. Aging & Mental Health 2009;3(4):497-511; doi:10.1080/13607860902774436
3. Lindsay J, Anderson L. Dementia/Alzheimer's Disease. Ottawa: University of Ottawa 2004
4. Alzheimer's Disease International. World Alzheimer Report 2009. London: Alzheimer's Disease International; 2009; www.alz.co.uk/research/files/WorldAlzheimerReport.pdf; retrieved April 27, 2010.
5. Ferri CP, Prince M, Brayne C, Brodaty H, Fratiglioni L, Ganguli M, Hall K, Hasegawa K, Hendrie H, Huang Y, Jorm A, Mathers C, Menezes P, Rimmer E, Scazufca M. Global prevalence of dementia: a Delphi consensus study. Lancet 2005;366(9503):2112-7; doi:10.1016/S0140-6736(05)67889-0
6. Steadman PL, Tremont G, Davis JD. Premorbid Relationship Satisfaction and Caregiver Burden in Dementia Caregivers. Journal of Geriatric Psychiatry and Neurology 2007;20(2):115-119; doi:10.1177/0891988706298624
7. Gaugler JE, Edwards AB, Femia EE, Zarit SH, Stephens M-AP, Townsend A, Greene R. Predictors of Institutionalization of Cognitively Impaired Elders: Family Help and the Timing of Placement. Journal of Gerontology; Series B: Psychological Sciences 2000;55B(4):247–255
8. Schulz R, Belle SH, Czaja SJ, McGinnis KA, Stevens A, Zhang S. Long-term care placement of dementia patients and caregiver health and well-being. JAMA 2004;292(8):961–967; doi:10.1001/jama.292.8.961
9. Komisar HL, Thomps LS. National Spending for Long-Term Care. Washington: Health Policy Institute, Georgetown University; 2007
10. Spector WD, Fleishman JA, Pezzin LE, Spillman BC. The Characteristics of Long-term Care Users. AHRQ Publication No. 00-0049. Rockville: Agency for Healthcare Research and Quality; 2001
11. Vredenburg K, Isensee S, Righi C. User-Centered Design: An Integrated Approach. Upper Saddle River: Prentice Hall; 2002
12. Feng J, Sears A, Law C. Conducting Empirical Experiments Involving Users with Spinal Cord Injuries. 3rd Interaction Conference on Universal Access in Human Computer Interaction (HCI); Las Vegas; 2005
13. Neuhauser L, Rothschild B, Graham C, Ivey SL, Konishi S. Participatory Design of Mass Health Communication in Three Languages for Seniors and People With Disabilities on Medicaid. American Journal of Public Health 2009;99(12): 2188-2195; doi:10.2105/AJPH.2008.155648
14. Harniss M, Amtmann D, Cook D, Johnson K. Considerations for Developing Interfaces for Collecting Patient-Reported Outcomes that Allow the Inclusion of Individuals with Disabilities. Medical Care 2007;45(5,Suppl 1):S48-S54; doi:10.1097/01.mlr.0000250822.41093.ca
15. Adlam T, Carey-Smith B, Evans N, Orpwood R, Boger J, Mihailidis A. Implementing Monitoring and Technological Interventions in Smart Homes for People with Dementia: Case Studies. In: Gottfried B, Aghajan H, editors. Behaviour Monitoring and Interpretation - BMI: Smart Environments. Amsterdam: IOS Press; 2009
16. Mihailidis A, Boger J, Craig T, Hoey J. The COACH prompting system to assist older adults with dementia through handwashing: An efficacy study. BMC Geriatrics 2008;8(28);doi:10.1186/1471-2318-8-28
17. Mihailidis A, Boger J, Canido M, Hoey J. The use of an intelligent prompting system for people with dementia: A case study. ACM Interactions 2007;14(4):34-37
18. Boger J, Hoey J, Poupart P, Boutilier C, Fernie G, Mihailidis A. A planning system based on Markov decision processes to guide people with dementia through activities of daily living. IEEE Transactions on Information Technology in Biomedicine 2006;10(2):323-333; doi:10.1109/TITB.2006.864480
19. Lenker JA, Nasarwanji MF, Paquet V, Feathers D. A tool for rapid assessment of product usability and universal design: Development and preliminary psychometric testing. Work: A Journal of Prevention, Assessment, and Rehabilitation 2010; in press
20. Law CM, Vanderheiden GC. Reducing sample sizes when user testing with people who have, and who are simulating disabilities – Experiences with blindness and public information kiosks. International Ergonomics Association and 44th Annual

Meeting of the Human Factors and Ergonomics Society; San Diego; 2000

21. Kurtz SM, Silverman J, Draper J. Teaching and learning communication skills in medicine. 2nd edition. Oxon: Radcliffe Publishing; 2005

22. Wethermon RA, Erbele S, Mattson M. Use of Standardized Patients as an Assessment Tool at the End of an Ambulatory Care Rotation. American Journal of Pharmaceutical Education 2000;64(Summer): 109-113

23. McNaughton N, Ravitz P, Wadell A, Hodges BD. Psychiatric Education and Simulation: A Review of the Literature. The Canadian journal of Psychiatry 2008;53(2):85-93

24. Folstein MF, Folstein SE, McHugh PR. Mini-Mental State: A practical method for grading the cognitive state of patients for the clinician. Journal of Psychiatric Research 1975;12(3):189-198; doi:10.1016/0022-3956(75)90026-6

25. Hoey J, Poupart P, von Bertoldi A, Craig T, Boutilier C, Mihailidis A. Automated Handwashing Assistance For Persons With Dementia Using Video And A Partially Observable Markov Decision Process. Computer Vision and Image Understanding - Special Issue on Computer Vision Systems 2010; 114(5); 503-519; doi:10.1016/j.cviu.2009.06.008

26. Rosen J, Mulsant BH, Bruce ML, Mittal V, Fox D. Actors' Portrayals of Depression to Test Interrater Reliability in Clinical Trials. American Journal of Psychiatry 2004;161(10):1909-1911; doi:10.1176/appi.ajp.161.10.1909

27. Hogarth RM, Einhorn HJ. Order effects in belief updating: The belief-adjustment model. Cognitive Psychology 1992;24(1):1-55; doi:10.1016/0010-0285(92)90002-J

28. Cunnington JP, Turnbull JM, Regehr G, Marriott M, Norman GR. The effect of presentation order in clinical decision making. Academic Medicine 1997;72(10):S40-S42; doi:10.1097/00001888-199710001-00014