Collaborative Educational Applications for Underserved Children: Experiences from India
Sumita Sharma

Collaborative Educational Applications for Underserved Children: Experiences from India

ACADEMIC DISSERTATION
To be presented with the permission of the Faculty of Communication Sciences of the University of Tampere, for public discussion in the Linna auditorium K103 on October 13th, 2018, at noon.

Faculty of Communication Sciences
University of Tampere

Dissertations in Interactive Technology, Number 29
Tampere 2018
**ACADEMIC DISSERTATION IN INTERACTIVE TECHNOLOGY**

| Supervisor: | Professor Markku Turunen, Ph.D.  
Faculty of Communication Sciences,  
University of Tampere,  
Finland |
|---|---|
| Opponent: | Associate Professor Payal Arora, Ph.D.  
Department of Media and Communication,  
Erasmus University Rotterdam,  
Netherlands. |
| Reviewers: | Professor Netta Iivari, Ph.D.  
Department of Information Processing Science,  
University of Oulu,  
Finland.  
Dr. Laura Malinverni,  
Department of Visual Arts and Design,  
University of Barcelona,  
Spain. |

Acta Electronica Universitatis Tamperensis; 1912  
ISBN 978-952-03-0807-0 (pdf)  
ISSN 1456-954X  
http://tampub.uta.fi

The originality of this thesis has been checked using the Turnitin Originality Check service in accordance with the quality management system of the University of Tampere.

**Dissertations in Interactive Technology, Number 29**

Faculty of Communication Sciences  
FIN-33014 University of Tampere  
FINLAND

ISBN 978-952-03-0808-7  
ISSN 1795-9489  
Juvenes Print – Suomen Yliopistopaino Oy  
Tampere 2018
Abstract

Interactive technology offers real world solutions for the learning needs of underserved children from emerging regions, including children with developmental disabilities and children from underprivileged backgrounds. Children with developmental disabilities face several social, cognitive, and learning challenges. Technology provides a safe, controlled, structured, and predictable environment and can be very appealing. Previous research has explored the potential and benefits of interactive technology, from gestures to full body interaction, for social, therapeutic, cognitive, motor, and life skills training. It is therefore no surprise that experts in Human Computer Interaction (HCI) and educational technology have designed, developed, deployed, and evaluated several applications for individuals with developmental disabilities, including Autism spectrum disorders and other learning disabilities. However, almost all of this activity is focused within the developed world and not within HCI for development (HCI4D).

Children from underprivileged regions face several socio-economic constraints regarding access to schools and teachers, computers and technology, relevant learning material, and the globalized world. Previous research has established the potential benefit of playful learning applications for out of school rural children and connecting them to the internet and to teachers from across the world. However, there is little research on globally inclusive collaboration in which school children from different socio-economic and cultural backgrounds connect using the same learning platform. Furthermore, cross-cultural collaborations and computer skills are considered important life and career skills to succeed in the 21st century. Thus, it is important to provide children from underprivileged regions a foothold into our global society.

In this dissertation, two research gaps within the domain of HCI4D for children are addressed. First, this work expands the currently limited research on interactive technology for children in developing regions by exploring ways to overcome the challenges towards introduction, access, and adoption. Second, previous research is primarily focused on individual interactions and not on collaboration. For children with developmental disabilities, this can further isolate them from their social group. For underprivileged children, this eliminates the opportunity to participate in the globalized world. To address these gaps, exploratory user studies with 107 children from two non-government organizations, Tamana and Deepalaya, in New Delhi were conducted. With Tamana,
gesture-based applications to promote social interaction and self-efficacy in children with developmental disabilities were designed, developed, and evaluated. With Deepalaya, user studies on online cross-cultural collaboration between India and Finland for English conversational language learning were conducted.

The key research takeaway is a set of guidelines for designing and evaluating collaborative applications for underserved children with a focus on reducing challenges towards technology introduction, acceptance, and adoption; designing research studies and conducting user evaluation; and designing interactive applications for children with developmental disabilities. This work provides fellow researchers, designers, and practitioners a way to approach designing and deploying technology for underserved children across the world.
Acknowledgements

This journey was made possible by the kindness, support, time and consideration of many remarkable people from across the world. I am grateful to my supervisor, Prof. Markku Turunen, who believed in me and told me I could do this, when I could not even think of it myself. I am also thankful to Prof. Netta Iivari and Dr. Laura Malinverni, for their valuable feedback, time, and interest in this research, and Associate Prof. Payal Arora, for graciously accepting to be my opponent in the Ph.D. defense.

The Tamana and Deepalaya organizations in New Delhi opened their doors to my research and accepted me as one of their own, for which I am grateful. I am also indebted to all the children who were part of this work; they made it truly memorable. It was an absolute pleasure and privilege to work with superb researchers from academia, industry, and non-government organizations. Nitendra Rajput, Saurabh Srivastava, Blessin Varkey, and Krishanveni Achary: what started as an internship made me realize my calling in life, I could not ask for more. Tomi Heimonen, Jaakko Hakulinen, Pekka Kallioniemi, Ville Mäkelä, Tuuli Keskinen, and Juhani Linna: thank you for all the wonderful and insightful discussions whenever I knocked on your doors – which was plenty!

I am thankful to my friends and family, who generously listened to me talk about my research and still be excited, the tenth time around. Mamma and Papa, you supported me unconditionally throughout this journey and everything before and after, I hope I made you proud. My darling sister, Neha Sharma, you are my rock.

Tampere, August 26th, 2018

Sumita Sharma
# Contents

1 **INTRODUCTION** ....................................................................................... 1  
  1.1 Objective .......................................................................................... 5  
  1.2 Context of Research ........................................................................... 7  
  1.3 Methodology ..................................................................................... 7  
  1.4 Results ............................................................................................... 8  
  1.5 Terminology ..................................................................................... 9  
  1.6 Structure of the Thesis ...................................................................... 12  

2 **GESTURE INTERACTION FOR CHILDREN WITH DEVELOPMENTAL DISABILITIES** .......... 13  
  2.1 The Learning Potential of Gesture-based Interaction ....................... 15  
  2.2 Gesture-Based Applications ............................................................... 16  
  2.3 Studies Employing Gesture Interaction for Indian Users .................... 23  
  2.4 Summary .......................................................................................... 26  

3 **COLLABORATIVE TECHNOLOGY FOR UNDERSERVED CHILDREN** .................... 29  
  3.1 Cross-Cultural Collaborative Learning ............................................... 31  
  3.2 The Bollywood Method ..................................................................... 32  
  3.3 Children, Computers, and Collaboration In India .............................. 33  
  3.4 Summary .......................................................................................... 36  

4 **INTRODUCTION TO THE RESEARCH CONTEXT, METHODOLOGY, AND APPLICATIONS** .... 37  
  4.1 The Research Context ....................................................................... 38  
  4.2 The Research Methodology ............................................................... 41  
  4.3 Balloons Application – Promoting Joint Attention ............................. 45  
  4.4 Kirana Application – Supporting Life-Skills .................................... 46  
  4.5 HOPE Application – Cognitive and Motor Skills Training ............... 48  
  4.6 CityCompass Application – Collaborative Language Learning ........ 50  

5 **SUMMARY OF THE PUBLICATIONS** .................................................................. 51  
  5.1 Promoting Joint Attention for Children With Autism ....................... 52  
  5.2 Gesture Interaction for Individuals with Developmental Disabilities ... 54  
  5.3 Kirana: A Gesture-based Market App for Life Skills ......................... 56  
  5.4 Guidelines for Designing Gesture-based Applications ..................... 57  
  5.5 The Bollywood Method for Cross-Cultural Collaboration .................. 59  

6 **GUIDELINES FOR DESIGN, DEVELOPMENT, AND EVALUATION OF APPLICATIONS FOR UNDERSERVED CHILDREN** .................................................... 63  
  6.1 Guidelines for the Underserved Context ........................................... 64  
  6.2 Guidelines for the Research Study Design ......................................... 66  
  6.3 Guidelines for Applications Design .................................................... 70  
  6.4 Summary of Guidelines ..................................................................... 75  
  6.5 Discussion .......................................................................................... 77  

7 **CONCLUSION** .......................................................................................... 85  

REFERENCES ................................................................................................. 87
List of Publications

This dissertation is composed of a summary and the following original publications, reproduced here by permission.


The Author’s Contribution to the Publications

The work presented in this dissertation was a collaboration between TAUCHI (Tampere Unit for Human Computer Interaction) in Finland and three Indian partners - IBM Research India, Tamana, and Deepalaya.

Publications I-IV

The goal was to develop novel gesture-based applications for children with developmental disabilities in India, together with TAUCHI, IBM Research India, and Tamana. My role included conducting user centered design studies with Saurabh Srivastava in New Delhi; designing, developing, and deploying gesture-based applications at two Tamana centers; and analyzing and publishing the results of the studies. Tomi Heimonen and Jaakko Hakulinen provided the programing framework used to develop the applications. Krishnaveni Achary and Blessin Varkey participated in the user centered design discussions and moderated the user studies to evaluate the applications with children with developmental disabilities. Nitendra Rajput and Markku Turunen mentored and managed the work. In addition, Srivastava, Hakulinen, and Heimonen contributed content for Publications I & II, while the other co-authors provided thorough reviews. Publication IV was thoroughly reviewed by Achary and Varkey. This work would not have been possible without the contribution and support of my colleagues at TAUCHI, IBM Research India, and Tamana.

Publication V

The user studies at Deepalaya utilized CityCompass, a collaborative language learning application based on a previously developed application called BerlinKompass. The pedagogical and language learning concepts of the applications were developed by Laura Pihkala-Posti, from the research center Plural. The applications were implemented by Pekka Kallioniemi and Jaakko Hakulinen, from TAUCHI. In India, Kallioniemi and I conducted user studies using CityCompass with children from Deepalaya, where I moderated the sessions in New Delhi and Kallioniemi connected from Tampere. I was also responsible for the setup in Delhi and analyzing and publishing the results of the studies. Additionally, all the co-authors assisted in the data analysis and thoroughly reviewed the paper. The generous support of my colleagues at TAUCHI made the work at Deepalaya possible.
# 1 Introduction

This dissertation presents exploratory design and user studies that establish the potential benefits of collaborative applications for children who are currently underserved by technology due to lack of informational, socio-cultural, or economical resources. This includes two specific groups of children – children with autism and other developmental disabilities and children from underprivileged backgrounds. In this work, developmental disabilities refers to neurodevelopmental disorders affecting cognitive and/or physical development of a child. This includes, Autism spectrum disorders (ASD), Down syndrome, and other unspecified forms of cognitive disorders co-morbid with attention deficit hyperactivity disorder (ADHD) or cerebral palsy (CP). Children from underprivileged socio-economic backgrounds, for the purpose of this work, refers to children living in informal urban settlements who may or may not attend public schooling. The children are underserved by technology due to informational, socio-cultural, infrastructural, and/or financial constraints. The work presented in this dissertation overlaps with several subfields within Human Computer Interaction (HCI) including computer supported collaborative work, interaction design and children, and accessibility. However, it is situated within the Human Computer Interaction for development (HCI4D) context to represent underserved or marginalized but emergent users, specifically focusing on users from India.

There are several challenges in introducing new technological interventions in an underserved context. From the lack of resources and limited access to infrastructure, electricity, and the internet to the existing digital divides between a diverse social strata (Agarwal, Sampath, & Indurkhya, 2012; Hajela, Bhattacharya, & Banerjee, 2013; Kam, Rudraraju, Tewari, & Canny, 2007; Kumar et al., 2010; Mitra, 2009). Furthermore, children with development disabilities face additional challenges in a
developing country’s context (Hajela et al., 2013). From low inclusivity and integration among children with disabilities and typically developed peers to stronger cultural barriers against children with disabilities that lead to pronounced digital exclusion, even within technology-capable households. For children from low-income households, access to technology has its costs: monetary, effort, and time. For instance, children can be working members of the family or their schools can charge an extra fee for computer classes (Kumar et al., 2010). There is also a growing aspiration among parents from underprivileged regions towards English and computer proficiency for their child in order to improve their future employment prospects (Pal, Lakshmanan, & Toyama, 2007). However, applications that tie in essential life and career skills – English, computer literacy, and cross-cultural collaborations (Trilling & Fadel, 2009) – are not well studied. Therefore, I focused on socially or globally collaborative applications for underserved children.

One potential way to mitigate technology adoption costs is through deploying interventions within community centers or schools. In this way, one device caters to a large group of children, and parents are able to witness the benefits for themselves before deciding to buy a device for individual use (Hajela et al., 2013). Moreover, within a school environment, technology can be used to empower educators as agents of change (Bhattacharya, Gelsomini, Pérez-Fuster, Abowd, & Rozga, 2015). Therefore, for this research, I worked extensively with two non-government organizations (NGOs) in New Delhi: Tamana1 and Deepalaya2.

The focus of the studies with Tamana was on collaborative gesture-based applications for children with developmental disabilities, while the focus with Deepalaya was on cross-cultural collaborative applications for children from underprivileged regions. The work presented in the dissertation is itself a cross-cultural collaboration between the research group, TAUCHI (Tampere, Finland), IBM Research India, Tamana, and Deepalaya (New Delhi, India). As is usual in HCI, the work is also interdisciplinary and includes experts from technology and (special) education. Each organization and its individuals brought in their own process, agendas, and aspirations towards the work. This required creating a rapport between the research organizations and the NGOs, as expected. As an Indian researcher, NGOs in my home city of New Delhi were accessible, and I was able to work with Tamana and Deepalaya to mediate the cross-cultural collaborations. Although the research location is Delhi, as the network of partners in Delhi provided access to underserved groups of children and experts working with them, the research outcomes are potentially transferable to underserved children in other locations across India and the Global South. However, further research and

1 www.tamana.org
2 www.deepalaya.org
resources are required to understand the nuanced differences in the culture and context of use.

**Collaborative gesture application for children with developmental disabilities**

Applications that employ gestures as their interaction mechanisms for imparting social, cognitive, and life skills provide a multitude of benefits to children with developmental disabilities. Children with autism are understood to lack motivation towards social interaction, face challenges in verbal and nonverbal methods of communication, and present several repetitive behaviors (DSM-5®, 2013). The degree of the challenges experienced depend on the individual, and autism is therefore understood on a spectrum. Interactive technology offers various advantages for individuals with developmental disabilities including controllable input stimuli, predictable and safe multisensory learning environments, and opportunities for customized and repeatable individualized learning goals (Bartoli, Corsardi, Garzotto, & Valoriani, 2013; Garzotto et al., 2014; Malinverni & Parés, 2017; Mora-Guııard, Crowell, Parés, & Heaton, 2016; Mora-Guııard et al., 2017; Parés et al., 2004; Parés, Masri, Wolferen, & Creed, 2005).

Research over the last several decades has steadily progressed in its application of technology for individuals with developmental disabilities with a specific focus on those on the autism spectrum. It dates back to the early 1970s, when speech based computer mediated interventions were designed for children on the autism spectrum (Colby, 1973). Since then, interaction mechanisms have become more immersive than the traditional mouse and keyboard interfaces and range from touch and eye-gaze to gestures and full-body interaction. The learning potential of gesture-based or full-body interaction is justified under the realms of embodied interaction, which itself is strongly based on the theories on embodied cognition with roots in psychology (Kirsh, 2013; Malinverni, 2016; Malinverni, Silva, & Parés, 2012). Owing to the success of Microsoft Kinect motion sensor for Windows, released in 2012, there was a surge in research focused on gesture-based interaction for individuals with developmental disabilities. As a part of this research, gesture interaction, including pointing at virtual objects and moving virtual objects using grab-drag-drop, was studied.

The gesture-based applications I designed and developed include *Balloons*, in which children with autism collaborate with a typically developed peer to select a virtual balloon together thereby experiencing shared attention, referred to as joint attention. *Kirana*[^3], in which individuals with developmental disabilities can practice the life skill of buying groceries from a local Indian store by going through the subtasks of decision making (deciding what to buy), arithmetic (calculating if they have

[^3]: A local grocery store in Delhi is called *kirana* in Hindi.
enough money to buy), and social interaction. HOPE improves motor coordination and social and cognitive skills, with increasing levels of difficulty, by providing tasks for matching colors, shapes, and objects. HOPE was designed and developed outside of the work presented in this dissertation. These gesture-based collaborative interventions are a first in India and supported by the pioneering spirit of Tamana. The work in this dissertation is presented with the same spirit – to create a starting point for researchers working with individuals with developmental disabilities within HCI4D.

Cross-cultural collaborations with children from underprivileged backgrounds

Cross-cultural collaborations, computer skills, and speaking English are all considered important skills to be a part of the global society of the 21st century (Trilling & Fadel, 2009). For children from underprivileged backgrounds, obtaining computer and English skills are understood to improve their future employment opportunities (Pal et al., 2007; Shukla, 1996), and therefore several researchers have focused on these two skills (Dangwal, Jha, Chatterjee, & Mitra, 2005; Kam et al., 2008; Kam, Ramachandran, Devanathan, Tewari, & Canny, 2007; Kam, Rudraraju, et al., 2007; Larson, Rajput, Singh, & Srivastava, 2013; Mitra, Tooley, Inamdar, & Dixon, 2003). In the ‘School in the Cloud’⁴, children in urban and rural India connect to a teacher from across the world using Skype. The main idea is to encourage problem-solving and peer collaborations using a computer and the internet with minimal supervision from the teacher. Most of the previous research with underprivileged Indian children is focused on children in rural schools or out-of-school children in rural areas. However, one-on-one cross-cultural online collaboration for school aged children in India is not well studied.

The Deepalaya organization in New Delhi educates underprivileged children primarily living in urban informal settlements. Using the CityCompass⁵ application, I ran a research study with children aged 11-12 years in Deepalaya in which they collaborated with a Finnish researcher. CityCompass allows cross-cultural collaboration between two remotely located users to complete a wayfinding task by navigating panoramic images of actual cities. I used the English version of the application and child-researcher pairs navigated the city of Tampere. In this dissertation, the challenges towards collaboration arising from socio-cultural norms and tech-inexperience are discussed and a method to overcome these challenges is suggested. For the educational and learning aspects of the application, which are out of scope for this dissertation, please refer to previous work with BerlinKompass (Pihkala-Posti, 2012 & 2014; Pihkala-Posti et al., 2013 & 2014; Kallioniemi, Pihkala-Posti et al., 2015).

⁴ https://www.theschoolinthecloud.org/
⁵ www.citycompass.sis.uta.fi
1.1 OBJECTIVE

There are two gaps in current research on HCI4D for children that this dissertation aims to fill. First, there is relatively little focus on technological interventions for underserved children in developing regions. For children with developmental disabilities, only two projects in India prior to this work are identified: one of which presents a short evaluation with only four participants with autism (Agarwal et al., 2012), while the other reports a framework but no evaluation (Hajela et al., 2013). To the best of my knowledge, the work with children with developmental disabilities presented in this dissertation is a first of its kind in India. For children from underprivileged regions, previous work is primarily focused on rural regions and on mobile applications for gamification of learning, especially for out of school rural children (Kam et al., 2008; Kam, Ramachandran, et al., 2007; Larson et al., 2013) or connecting children to educators across the world for minimally supervised online learning (Dangwal et al., 2005; Dolan et al., 2013). There is limited research on children from urban underprivileged regions collaborating online with a developed world adult or peer.

Second, previous research is primarily focused towards individual interactions. Children with developmental disabilities face challenges in social interactions and, therefore, interventions made for single person interactions exacerbate isolating experiences. In my research, I focused on socially collaborative applications using gestures for interaction with co-located peers. This is similar to previous work in Milan by Mora-Guiald (2016). Children from underprivileged regions are encouraged to collaborate with their co-located peers (Mitra, 2009; Mitra & Rana, 2001) but not towards online cross-cultural collaborative learning and experience, which are important career skills (Trilling & Fadel, 2009). These research gaps aggravate the growing digital divide for currently underserved children in the Global South, which, in turn, attenuates the disparity between the haves and have-nots and therefore require attention. With partners and resources available in New Delhi, my work on understanding how to bridge these gaps started in India.

The overall goal of this work is to answer the research question:

RQ: How to design and introduce collaborative educational applications for social, life, and career skills to children who are currently underserved by technology due to lack of informational, socio-cultural, or economical resources?

This question is divided into three sub-questions, SQ1-3, which are addressed by conducting exploratory user studies with three different user groups. The sub-questions are:

---

6 Using google search and google scholar with the keywords: Autism Kinect India, Kinect India, and Gesture HCI India.
SQ1. Children with autism: How to promote social collaboration between children with medium-low functioning autism who lack motivation for social interactions and a typically developed peer (e.g., sibling, care-giver, school-mate)?

SQ2. Children with developmental disabilities: How to support the practicing of a life skill for children with developmental disabilities in a safe and controlled environment which also translates to real world scenarios?

SQ3. Children from urban informal settlements: How to encourage cross-cultural collaborations between underprivileged children and a developed world peer?

The sub-questions are answered through a series of exploratory user studies with individuals studying in three schools associated with the two NGOs, Tamana, and Deepalaya, using four applications. The user studies are presented in the publications that are a part of this dissertation. Table 1 presents an overview of the research discussed in this dissertation.

<table>
<thead>
<tr>
<th>Research sub-question</th>
<th>SQ1</th>
<th>SQ2</th>
<th>SQ3</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Demographics</td>
<td>Children with autism</td>
<td>Individuals* with developmental disabilities</td>
<td>Children from urban informal settlements</td>
</tr>
<tr>
<td>School or Center</td>
<td>Tamana School of Hope</td>
<td>Nai Disha Vocational Education Center (Tamana)</td>
<td>Deepalaya Learning Center Sanjay Colony</td>
</tr>
<tr>
<td>Focus of the study</td>
<td>Social interaction</td>
<td>Life skills</td>
<td>Cognitive skills</td>
</tr>
<tr>
<td>Applications</td>
<td>Balloons</td>
<td>Kirana</td>
<td>Balloons and HOPE</td>
</tr>
<tr>
<td>Number of Participants</td>
<td>10</td>
<td>36</td>
<td>14</td>
</tr>
<tr>
<td>Publications</td>
<td>I</td>
<td>II, III, and partially IV</td>
<td>IV</td>
</tr>
<tr>
<td>Guidelines</td>
<td>Sections 6.1, 6.2, and 6.3</td>
<td>6.1 and 6.2</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Overview of the research presented in the dissertation
(*several participants are over the chronological age of 18 years; **collaboration was between child-researcher pairs)

Based on the results of exploratory research studies with the different groups, this dissertation provides an exhaustive list of guidelines for fellow researchers, designers, and practitioners working with technology for underserved children on how to design, deploy, and evaluate similar applications within a school environment.
1.2 CONTEXT OF RESEARCH

The context of this research is underserved children. Specifically, it focuses on children in Tamana (children with autism and other developmental disabilities) and Deepalaya (children from underprivileged urban regions) and the challenges they face towards technology access and adoption. For instance, this includes the socio-technical aspirations of parents for their children and the experience of autism in India. Deepalaya and Tamana also share several similar traits: their building architecture and its openness, and as NGOs focused on underserved children and creating an impact on their lives through education and training.

Tamana has catered to individuals with developmental disabilities and has been a pioneer in special education since its first center opened in 1985 in New Delhi. Tamana has four centers or special schools, of which I worked with two - the Tamana School of Hope and Nai Disha Vocational Education Centre. The School of Hope caters to children with medium-low functioning autism and Nai Disha caters to individuals with developmental disabilities including children with high-functioning autism. Two researchers, one at each center, worked on designing, deploying, and evaluating the applications that were developed during this work.

Deepalaya caters to children from low socio-economic backgrounds and has four centers in Delhi and another three in the neighboring states. I worked with the Deepalaya Learning Center Sanjay Colony in Okhala (one of Delhi’s largest industrial areas). At Deepalaya Learning Center Sanjay Colony, which from here on is referred to as Deepalaya, children are provided remedial education to supplement their public-school classes, which are considered to be inadequate. Children from Deepalaya interacted with a Finnish researcher from TAUCHI in Tampere to complete collaborative tasks using CityCompass.

1.3 METHODOLOGY

The user studies presented in this dissertation are exploratory in nature and are similar to ethnographical methods largely used in HCI4D. Both quantitative and qualitative data were collected during the studies to substantiate the potential of the applications and formulate a set of guidelines for further research in this domain. The work presented in this dissertation has a high ecological validity as it was evaluated in the field – classrooms at Tamana and Deepalaya instead of a university or laboratory setting.

The applications developed for children with autism and other developmental disabilities followed a user-centered design process. At the Tamana School of Hope, this included special educators, therapists,
parents, and two researchers at Tamana who are responsible for all technology mediated learning. Together with a researcher from IBM Research India, I spent time at the school familiarizing myself with the environment and its people, and then I spent the next two weeks interviewing the participants in small focus groups. At the end of the process, I gathered a set of guidelines which led to the development of the Balloons application. Balloons promotes social interaction between a child with autism and a typically developed peer or adult using gestures for interaction.

Kirana is a gesture-based application that supports the life skill of buying groceries from a local store. Kirana was developed after an extensive user-centered design study conducted at Nai Dish with children with developmental disabilities and their parents during an annual school fair, called Diwali Mela, to introduce gesture-based interaction. A gesture vocabulary for children with developmental disabilities was also identified. The HOPE application was developed by Tamana, after studies with Balloons and Kirana, and I was involved in its evaluation with children with developmental disabilities at Nai Dish. The work with Tamana includes 60 children with autism and other developmental disabilities.

I also conducted two exploratory user studies with 47 children from Deepalaya. From the findings of the first study, it became evident that participants faced several socio-cultural barriers towards communication with a Finnish researcher, which in turn affected their interaction with CityCompass and, therefore, the overall collaboration. The Bollywood Method, in which tasks are rooted in dramatized scenarios, is known to overcome similar socio-cultural barriers to communication as previously tested with Indian adults (Chavan, 2005; Chavan, Gorney, Prabhu, & Arora, 2009). Therefore, in the second study, the Bollywood Method was adopted for cross-cultural collaborations for Indian children.

1.4 Results

The sub-questions are answered through a series of exploratory user studies with individuals studying in three schools associated with the two NGOs, Tamana, and Deepalaya, using four applications. The user studies are presented in the publications that are a part of this dissertation. The key research takeaway from this work is a set of guidelines for:

(a) introducing interactive technology to underserved children in developing regions (section 6.1),
(b) addressing issues towards the research study design, especially when conducting user studies with underserved children in developing regions (section 6.2), and

(c) designing interactive applications for children with developmental disabilities in India (section 6.3).

Guidelines in (a) and (b) are extendable to underserved children both in developing and developed country contexts, that is, the urban-poor, as the resource constraints towards technology access can be monetary, informational, infrastructural, or socio-cultural. Guidelines in (c) were derived from the experience with user studies at Tamana with gesture-based interactions, but they are also extendable to other interaction mechanisms. Furthermore, several of the guidelines in (c) are seemingly universal for children with autism across the world. The applications developed for Tamana, Balloons, and Kirana, are also results of my research.

1.5 TERMINOLOGY

Within the scope of the research presented in this dissertation, several common terms are used in a specific way which are clarified here.

The Underserved Context

Being underserved is the result of a complex composition of various factors, including a lack of access to infrastructure, e.g. roads, school-buildings, and basic public services, e.g. healthcare, education. However, for the purpose of the work presented in this dissertation, the term underserved is used to denote an environment that suffers from a lack of informational, socio-cultural, or economical resources, which, in turn, affects access to technology understood to be beneficial. Therefore, children in an underserved context do not have access to technology that is known to potentially benefit their learning of social, life, and career skills, in addition to an already constrained surrounding. This also includes a lack of access to devices, such as laptops, Microsoft Kinect sensors, and culturally-appropriate games or applications.

Technology Access

Technology in developing regions exists is many different states of use and availability depending on infrastructural, informational, socio-cultural, and economical perspectives. For instance, even within urban environments which support infrastructural access, e.g. New Delhi has 4G mobile networks, and an availability of affordable smart phones with a cheap data connection, this still might not “translate into effective use” (Gitau, Marsden, & Donner, 2010). Therefore, in addition to access from an infrastructural and economic perspective, there ought to be informational
resources on how to start socio-cultural approval and an awareness of the benefits of using a specific technology. Thus, access to technology is truly achieved when there is an overlap of social and individual motivation towards its usage. In this dissertation, the lack of technology access is understood to be a consequence of limited informational, socio-cultural, or economical resources.

**Technology Introduction, Acceptance, and Adoption**

The term technology introduction is used to refer to the way a specific technology is first introduced within a specific context of use where the stakeholders involved have limited or no previous experience with that technology. For instance, Balloons, an application using gesture-based interaction, was designed and developed for a special needs classroom using a user-centered process. The different stakeholders involved (parents, teachers, and children with developmental disabilities) had no previous experience with gesture-based interactions. Therefore, it was important to study gestures, using a Kinect, to be accepted by the different stakeholders as a potential medium of interaction. Technology acceptance also plays a crucial role after the technology is introduced. It determines long term technology adoption, that is, whether the applications developed still be in use after the user study has ended and researchers have left. The research presented in this dissertation focuses on challenges towards technology introduction and acceptance and extrapolates their influence towards adoption.

**HCI4D: Developing vs Developed Countries**

Given the strides made in the age of information and globalization, it is quite understandable that the line between developing and developed, within the scope of an entire country, is blurry. Add to this the scale of the Indian populace, in which many realities are juxtaposed within the ever increasing boundaries of a city. However, for the sake of drawing comparisons and parallels between the context of a city like London, Milan, or Barcelona, when compared to New Delhi, the developed and developing world terminology is used.

**Autism Spectrum Disorders**

The term autism spectrum disorders (ASD) is used to represent varying levels of behavioral, cognitive, and social challenges an individual with the diagnosis faces (DSM-5®, 2013). Individuals with autism have reduced motivation towards social interactions and may face challenges towards verbal and non-verbal communication. They are also known to display repetitive behaviors. Since the severity and type of challenges experienced vary greatly among individuals with autism, the terms low, medium, or high functioning autism are used as an informal sub-classification.

In the user studies presented in this dissertation, participants were classified as either medium functioning individuals with autism or low
functioning individuals with autism. Individuals with medium-high functioning autism communicated with a vocabulary of 150–500 words, had an IQ of over 50, could follow instructions provided to them, and had an awareness of self and body. Whereas, individuals with low functioning autism were at the severe end of the developmental disability spectrum. They had an IQ of less than 50, communicated with a vocabulary of 50–150 words, could not follow instructions, but had an awareness of self and body.

Children with Developmental Disabilities
The term developmental disabilities is used to refer to a wide range of neuro-developmental and/or physical disabilities including (but not limited to) autism spectrum disorders, Down syndrome, and other learning disabilities co-morbid with ADHD, cerebral palsy, and other motor impairments (DSM-5®, 2013). The words children and individual are also used interchangeably to refer to both children and adults, which, in terms of chronological ages, includes those between 8 to 40 years. This is because chronological age is not an ideal representation of the abilities of an individual with developmental disabilities. Section 4.2 discusses the parameters used for recruiting participants for a study and the average age groups of each study.

Gesture-based Interaction
In HCI, a gesture is defined as any physical movement of the body including head, hands, and feet, that can be recognized by a system as an input without the need for any external hand-held devices (Saffer, 2008). With this definition, waving, nodding, finger-swipes on a mobile screen, and contouring facial muscles can be a gesture. However, for the context of this work, gesture interaction mainly refers to midair hand movements which include pointing, movement of the hands and arms, and closing or opening of the palms. This allows for interactions using gestures such as pointing with a dwell time and grab-drag-drop.

Collaboration vs Cooperation
From the perspective of the work with children from Deepalaya, two people are required to complete the tasks in CityCompass, and thus, the word collaboration is used. However, in literature on embodied interaction for individuals with developmental disabilities, the word collaboration is used quite leniently to refer to any activity or task that requires more than one person or any interaction among two people, with or without actual interaction with technology by both people (Bernardini, Porayska-Pomsta, & Smith, 2014; Mora-Guiard, Crowell, Parés, & Heaton, 2017; Parés, Carreras, & Durany, 2004). In Balloons, two people are required to interact with virtual elements to yield a bigger virtual reward, thus indicating collaboration in the strictest of sense. In order to keep the terminology simple, I also use the word collaboration when discussing
Kirana and HOPE, in which the social interaction occurs with a peer or an adult moderator while only the participant interacts with the application. This is usually referred to as co-operation within the domain of HCI. Furthermore, during the evaluations with Balloons, moderators first physically guided the pointing gestures, as shown in Figure 1. In this case, the system identified them as one user and, therefore, the reward was smaller, but this type of interaction is also considered collaborative.

![Figure 1: A moderator and participant collaborate to execute the pointing gesture](image)

### 1.6 Structure of the Thesis

This dissertation is a collection of publications with a summary. The summary presents an overview of related work, the research context, the applications, and the publications. Sections 2 and 3 present the related work in the research domain including the challenges faced by individuals with developmental disabilities in India and by children from underprivileged regions towards educational technology access and usage. Section 4 introduces the research context and describes the applications that were developed and evaluated during this research. Section 5 summarizes the publications. Section 6 presents an exhaustive set of design guidelines based on the research conducted for this work. The guidelines are grouped into three categories: introducing technology to an underserved context, considerations for conducting a user study, and designing gesture-based applications specifically for children with developmental disabilities. The applications and guidelines are the main contribution of this work (Section 4 and 6). The dissertation ends with a short discussion on the work and presents a vision for future work in this domain.
2 Gesture Interaction for Children with Developmental Disabilities

Gestures are an important part of human communication and social interactions. They are a “robust phenomenon found across culture, ages, and tasks” (Goldin-Meadow, 1999). Gestures are also natural and intuitive; individuals who are blind since birth use gestures that are similar to their sighted peers, although in lesser frequency (Iverson & Goldin-Meadow, 2005). It is therefore no surprise that emerging interactive technologies incorporate gestures for interaction (Bhuiyan & Picking, 2009). With the introduction of the Microsoft Kinect sensor for Xbox 360 in 2010 and the Wii controller before it, the landscape of gaming changed. The Kinect motion sensor for the Windows operating system was first made available in 2012 in an attempt to broaden Xbox 360's audience beyond its gamer base. As expected, this opened up exciting opportunities for researchers and designers to develop prototypes, applications, and games to study gestures for interaction. This includes research into public displays, of which the focus ranges from understanding user attitude and attention towards interactive public displays (Kukka, Oja, & Kostakos, 2013; Müller et al., 2009), analyzing a user’s in the wild interaction experience (Kellar et al., 2005; Keskinen et al., 2013; Peltonen et al., 2008) to examining audience or passer-by behavior and the inherent social dynamics with respect to user interaction (Brignull & Rogers, 2003; Michelis & Müller, 2011; Perry, Beckett, O’Hara, & Subramanian, 2010).

Research is also increasingly moving away from laboratory environments to other indoor and outdoor social locations. These locations can include museums (Hall & Bannon, 2005; Keskinen et al., 2013), control rooms
(Heimonen et al., 2016), housing fairs (Sharma, 2013), shop windows (Peltonen et al., 2008), and universities (Mäkelä, Sharma, Hakulinen, Heimonen, & Turunen, 2017; Müller, Bailly, Bossuyt, & Hillgren, 2014; Valkanova, Walter, Moere, & Müller, 2014). In this work, the focus is on a school environment.

In the Global North, research on gesture-based interaction for individuals with developmental disabilities has steadily increased over the last 15 years under the umbrella of embodied or full body interaction. Full-body interaction is defined as “the use of motion and activity in a physical space by the user’s body as the interaction mechanism” (Mora-Guiard et al., 2017). Several groups are dedicated towards designing, developing, evaluating, and deploying interactive systems for individuals with autism and other developmental disabilities. The research focus is mainly on improving a specific or combination of skills related to social, cognitive, motor, and life skills. The interaction varies from full-body gestures (jumping or running) and hand gestures (pointing or dragging and dropping), to a combination of embodied and tangible interactions (Bartoli et al., 2014; Malinverni et al., 2017; Mora-Guiard et al., 2017; Parés et al., 2005). The earliest project documented in literature to use full-body gestures for children with autism is MEDIATE (2002 – 2012) (Parés et al., 2005, 2004; Parés, Soler, et al., 2005; Parés, Masri, Van Wolferen, & Creed, 2005). The interactions were designed to take in inputs from pressure sensors on the walls and floors of the environment and through infrared cameras and lights. This project started in the pre-Kinect era, but employed similar motion tracking techniques.

Previous research has firmly established the benefits of interactive technology for individuals with developmental disabilities: a controllable input stimuli and predictable outcomes, a safe environment, the possibility of repetition, the possibility for multi-sensory output, customization for each individual, scalability to serve several individuals, and the potential to mediate social interactions, to name a few (Bartoli et al., 2013, 2014; Malinverni et al., 2012; Mora-Guiard, 2016; Mora-Guiard et al., 2017). Researchers have also observed an affinity towards technology-based solutions from individuals with developmental disabilities (Goldsmith & LeBlanc, 2004; Mineo, Ziegler, Gill, & Salkin, 2009). The work presented in this dissertation contributes to previous research in two ways: as a first look into the challenges towards technology introduction in an underserved context and, subsequently, as a set of guidelines for other researchers targeting similar regions.

In this section, I first discuss the potential of gestures for learning, especially of complex and abstract concepts. Then I present evaluations of applications that employ the Microsoft Kinect sensor for gesture recognition and an overview of the research on full-body interaction with
Given that the work presented in this dissertation is conducted in India and focuses on an underserved context, I also present studies that employed gesture interaction for underserved Indian users.

2.1 The Learning Potential of Gesture-based Interaction

Within HCI, the methods of interaction with a system have evolved from the traditional mouse and keyboard to the realm of using one’s own body. From sitting and staring at a screen to jumping or pointing in Kinect games, this interaction with systems creates an embodied perception of the body, mind, and technology (Kirsh, 2013). “The essential feature of embodied interaction is the idea [...] of allowing users to negotiate and evolve systems of practice and meaning in the course of their interaction with information systems” (Dourish, 2004). Embodied cognition theory, from neuroscience and cognitive psychology, states that cognition is situated, that is, “cognitive processes are deeply rooted in the body’s interactions with the world” (Wilson, 2002). As it is now understood, we think not just with our minds but also with our body. This informs us of the potential of embodied interaction to contribute to the learning experience of new content in which we use our bodies to interact with digital systems the way we would in the physical world. Embodiment can be achieved through multiple modalities and is an umbrella term that includes gesture-interaction using hands or the upper body and full-body motion. This includes granular movement in spatial locations including gesturing using hands, like pointing, and full-body motions like jumping or forming specific shapes using the body.

Within the education domain, a strong case of gesture interaction within a classroom environment is also supported by Gardner’s theory of multiple intelligences (Gardner, 1989), in which learners and the process of learning itself is classified into six categories, one of them being kinesthetic. Kinesthetic learning or learning by doing, is again supported by the theory of embodied cognition. Applications that employ gesture-based or full-body interaction to teach a specific life skill to children with developmental disabilities have the potential to support learners who benefit from kinesthetic approaches [Publications I & II]. Furthermore, research has shown the benefits of using embodied interaction for children with developmental disabilities for the process of learning complex and abstract concepts, such as those in mathematics (Abrahamson & Trninic, 2011; Okkonen, Sharma, Raisamo, & Turunen, 2016; Roth & Thom, 2009), and the concept of self (Bartoli et al., 2014; Parés et al., 2005).

The process of learning is complex and dependent on several factors in addition to the medium of interaction, such as the context or learning environment, the level of engagement with the content, and the design of
the educational application. Gesture-based interaction potentially scaffolds the process of learning for children with developmental disabilities (Malinverni, 2016) given that other factors also support learning. This dissertation focuses on supporting the learning of social and life skills for individuals with developmental disabilities using gestures as the medium of interaction. The design of applications and their elements, through a user-centered process [Publications I & II], also contributes to the learning experience. Next, previous research on applications using the gesture-based interaction, using upper body and full body motion, for children with developmental disabilities is discussed.

2.2 Gesture-Based Applications

One of the main goals of gesture-based applications for individuals with developmental disabilities is to promote social interaction. This includes initiating interaction with a peer, joint or shared attention towards the same object, turn-taking and collaborative problem solving, verbal or non-verbal communication, (Bartoli et al., 2014; Garzotto, Gelsomini, Oliveto, & Valoriani, 2014; Garzotto, Valoriani, et al., 2014), and sometimes also competition (Bhattacharya et al., 2015). Other goals include cognitive skill training for improving attention, memory, and object matching; academic skills such as arithmetic; and motor skills to improve posture stability, hand-eye coordination, and fine motor movements, which are often times inherent in the gesture mechanism for interaction (Altanis, Boloudakis, Retalis, & Nikou, 2013; Caro, Tentori, Martinez-Garcia, & Zavala-Ibarra, 2017; Kourakli et al., 2017). Overall, applications can be designed to encourage social, cognitive, academic, life skills, or motor goals. The applications presented in this dissertation used the Microsoft Kinect sensor to track an individual’s hand gestures, similar to the applications described next. The novelty of my work lies in its underserved context of use and its focus on socially collaborative applications.

M4ALL\footnote{http://www.m4allproject.eu/}, or Motion-based adaptable playful learning experiences for children with motor and intellectual disabilities, was an EU project with four partner universities, one non-profit, and one children’s clinic. The project was spread across five countries: Italy, Spain, Netherlands, Belgium, and Greece, and was active from 2012 till 2014. The applications developed as a part of the M4ALL employed the Microsoft Kinect motion sensor and provided customization of learning goals to enable therapists, parents, and caregivers to create individualized interventions (Altanis et al., 2013; Bartoli et al., 2014; Kourakli et al., 2017; Malinverni et al., 2017; Mora-Guiard, 2016). This includes Pico’s Adventure, a dual player game with a fantasy world in which Pico, an alien, has landed on earth and now needs help to fix his spaceship to return home (Malinverni et al., 2017;
Mora-Guiard, 2016). The goal of the game is to promote social interaction between individuals with autism and typically developed individuals, adults, or peers. An image of the child, captured by the Kinect camera, is superimposed on their virtual avatar within Pico’s virtual world. However, seeing their face on the virtual avatar was later found to be uncomfortable for the child (Mora-Guiard, 2016). Using hand gestures, like stretching, a child must reach objects within the virtual world to help Pico (as shown in Figure 2, right). The game presents a proof of concept and the potential benefits of co-located social interaction interventions to elicit social initiation by children with autism. For example, certain tasks, like collecting stars from the sky by holding hands, require the assistance of an adult or a peer to complete (Figure 2, left).

Evaluations of Pico’s Adventure were conducted with 15 children with medium-high functioning autism, aged between 4-7 years, at a hospital in Barcelona, during four hour-long sessions over the course of two months. In the first session, the child played alone, in the second and third sessions the child played with their parent, and in the last session, the child played with a peer. Overall, results show that Pico’s Adventure promoted social initiation in participants more effectively than in free play without parents, and, therefore, could potentially facilitate interactions between a child with autism and an unfamiliar individual. This result supports the potential of the Balloons application [Publications I & IV] to support social interaction, by joint attention, between a child with autism and a typically developed individual.

![Figure 2: Pico’s Adventure (left) with a parent and child collecting stars and (right) a child performing gestures individually](Malinverni et al., 2017) © Elsevier

Bubble, Space, and Shape games (Polimi games) were developed after a study with Xbox games for children with autism in Milan, which was also a part of the M4ALL project (Bartoli et al., 2014; Garzotto, Valoriani, et al., 2014). In the Bubble game, participants have to catch as many bubbles as possible within a given time using their onscreen avatar (stick figure). This helps with hand-eye coordination and self-realization or learning the concept of self. The participant is rewarded with applause and a golden cup if they catch a certain number of bubbles. The bubbles are caught with the head, hands, or feet. The Space game includes several virtual objects that fall from the top of the screen to the bottom and must be avoided by
the onscreen avatar of the participant. This is achieved by moving the whole body or the hand horizontally. This game uses the concepts of lives in a time bound session. In the Shape game, the focus is on body posture or awareness and imitation, where the participant has to replicate a given body shape with their virtual avatar. This can also be achieved by two children using their body for interaction (controlling their stick figure avatar), by holding a posture for a certain amount of time. All the games can be customized by the therapist for individualized learning goals and controlling the game elements (for example, the speed of the falling object in the Space game).

Polimi games were evaluated with 10 children with autism aged between 6-8 years at a therapeutic center in Milan over a period of three months (dividing the students into control and treatment groups). The findings provide strong empirical evidence on the benefits of touchless motion-based games (gestures and full-body), to promote attention and the integration of motor and visual skills. These findings formed the basis of my research with individuals with developmental disabilities in India (including autism) presented in this dissertation. I further expand the work to real world scenarios where selection via pointing and moving of physical objects simulates interactions in everyday life [Publications II & IV]. Moreover, guidelines derived from the results of the studies with Xbox games and Polimi games (Bartoli et al., 2014) further support the guidelines presented in section 6.3. This leads to the conclusion that several of the challenges faced by children with medium-low functioning autism, especially towards social interaction, are inherently universal, as also discussed in section 5.1 [Publication I].

Kinems.com, a company based in the US, has developed an entire suite of applications for children with special needs. The effectiveness of the games and potential benefits to children within the classroom and home environment are evaluated with partner schools across the US and Europe. The games focus on cognitive, motor, and academic skills and employ several gestures (pointing with dwell time for object selection and movement, jumping, grabbing and dropping) as shown in their product video. The games include Mathloons, in which users select the correct balloon out of three options to solve a mathematical task; Farm Walks, in which an avatar of a farmer is guided through a pathway of different shapes using hand gestures; Space Motif, in which users arrange planets and other objects in a given pattern without colliding with the neighboring objects; UnBoxIt, in which a flash card like approach is used to remember objects hidden in boxes; Melody Tree, in which the task is similar to UnBoxIt but, instead of objects, the boxes contain sounds; River Crossing, in which the user drags a boat across the river to transport

---

8 www.kinems.com
9 https://youtu.be/3mrrZlumIh0
animals and things; and Uni Paca Girl, in which the participant controls an avatar of a girl and guides her along a maze, avoiding obstacles and collecting rewards using hand gestures.

A study with over 20 children with ADHD and other learning disabilities such as dyslexia, aged between 6-11 years, using five of the Kinems games that employ hand gestures, showed positive improvements in their cognitive skills: short term memory, visual processing, and motor and mathematical skills (Kourakli et al., 2017). The study was conducted within an inclusive school that integrates typically developed and special needs children. However, the evaluations were conducted as individual interactive sessions with a special needs child and an educator or a moderator who assisted and observed. The Kinems application suite also incorporates an in-depth progress reporting system for each user for both learning and kinetic analysis. In another study, ten children with motor impairment and co-morbid learning disabilities, who used two of the Kinems games (Farm Walks and River Crossing), showed an improvement in motor skills with respect to the “precision, coordination, or strength” of executing a gesture, over a five month period (Kosmas, Ioannou, & Retalis, 2017). In a pilot study with Uni Paca Girl, two children with motor impairments used the game with their therapists and found it engaging and therapeutic (Altanis et al., 2013).

The Pictogram Room consists of several playful activities focused towards self-awareness and social interaction, including cooperation and collaboration, for an educator and a child with autism (Casas, Herrera, Coma, & Fernández, 2012; Herrera, Casas, Sevilla, Rosa, & Pardo, 2012). The educator chooses the activity and then instructs or collaborates with the child. The system augments reality with virtual objects and virtual puppets of the participants. Activities include learning about a body part and identifying themselves as different from another participant in the augmented reality scenario. Another activity involved creating certain body postures with the virtual puppet. Studies were first conducted with typically developed children and then with five children with autism. The level of assistance required by the child (verbal or physical) was also noted in the assessments. Figure 3 (left) shows a pictogram of a person drinking water to represent the action of drinking water and a participant mirroring that action. The participant’s online puppet is similar to the pictogram image, conceptualized as an augmented mirror. This augmented mirror helps children with autism to generalize tasks and actions, such as relating the pictogram to an actual action.

MEBook teaches social initiation by placing the participant in a virtual environment with other characters (Figure 3, right). The participant’s face is super-imposed on a virtual avatar. The game follows a narrative based approach in which the participant meets and greets different characters
and acts out the physical aspects of the greetings, such as waving. The application is designed to identify and respond only to socially acceptable greetings and ignores others (Uzuegbunam, Wong, Cheung, & Ruble, 2015).

**Figure 3**: (left) A pictogram of the action to drink water [(Herrera et al., 2012) © IEEE] and (right) MEBook environment [(Uzuegbunam et al., 2015) © IEEE]

ICAN\(^\text{10}\), formerly known as the Lakeside Center for Autism, developed several Kinect-based applications focusing on social, cognitive, motor, and speech skills for children with autism. Studies have also employed Kinect sensors to detect repetitive behaviors or tantrums in children with autism (Firth, 2012; Goncalves, Costa, Rodrigues, & Soares, 2014; Yu, Wu, Liu, & Zhou, 2011), develop emotion recognition in preschoolers with autism (Christinaki, Triantafylidis, & Vidakis, 2013), and improve motor skills for children with developmental disabilities (Bravo, Ojeda-Castelo, & Piedra-Fernandez, 2017; Caro et al., 2017; Da Gama, Fallavollita, Teichrieb, & Navab, 2015).

The work presented in this dissertation focuses on interaction using hand gestures for learning interventions for social, cognitive, and life skills development for children with autism and other developmental disabilities. It is similar to previous work described in this section and summarized Table 2. This includes Kinect applications that have been evaluated with individuals with developmental disabilities within a special school and other similar settings. One thing to note is the limited number of participants for evaluations of the applications, which limits the generalizability of the results. This is common when working with children with special needs. Furthermore, Pictogram Rooms and MEBook superimposed the face of the participant on their virtual avatar, which might have created uncomfortable experiences (Mora-Guiard, 2016). In my work with Tamana, onscreen hand cursors were used to guide the gesture interactions and the user studies were conducted with 10-18 participants, in line with previous studies in this domain.

\(^\text{10}\) http://i-can.center/
<table>
<thead>
<tr>
<th>Application</th>
<th>Goals</th>
<th>Interaction Mechanism</th>
<th>Evaluation Location</th>
<th>Evaluation Demographics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pico’s Adventure</td>
<td>Initiate Social Interaction, Motor Skills</td>
<td>Hand gestures, holding hands, body postures</td>
<td>Hospital, Barcelona</td>
<td>5-6-year-olds with medium/medium-high functional autism (15 participants).</td>
</tr>
<tr>
<td>Polimi games</td>
<td>Attention and motor skills</td>
<td>Using body postures, hand gestures</td>
<td>Therapeutic center, Milan</td>
<td>6-8-year-olds with autism (10 participants)</td>
</tr>
<tr>
<td>Kinems Games</td>
<td>Cognitive, Academic, and Motor skills</td>
<td>Hand gestures</td>
<td>Piraeus, London, and a Special School, Attica</td>
<td>6-11-year-olds with ADHD and other learning disabilities or motor impairments (Multiple studies: 32+ participants)</td>
</tr>
<tr>
<td>Pictogram Room</td>
<td>Self-awareness and social interaction</td>
<td>Body postures</td>
<td>Integrated school</td>
<td>Children with autism (5 participants)</td>
</tr>
<tr>
<td>MEBook</td>
<td>Social interactions: greetings/ initiation</td>
<td>Gestures for greeting (saying Hi, Bye, e.g. waving)</td>
<td>Not mentioned</td>
<td>7-12 year-olds with autism (3 participants)</td>
</tr>
</tbody>
</table>

**Table 2:** Summary of applications employing gestures-based interactions using the Kinect sensor and evaluated with individuals with developmental disabilities

In addition to Kinect-based applications, full body interactions employing other devices have also been explored for individuals with developmental disabilities. MEDIATE is a Multisensory Environment Design for an Interface between Autistic and Typical Expressiveness. It is one of the first notable projects that introduced full-body interaction, with other modalities, to a virtual environment (Parés et al., 2005, 2004; Parés, Soler, et al., 2005). During the design of the environment, potential full-body gestures were explored in relation to participants’ concept of the self and of being in control. That is, if objects on a screen reacted by growing in size or changing color to participant movements around the screen, the participant “would hopefully understand she is in control of the situation” (Parés, Soler, et al., 2005). The interactive environment used pressure sensors in the walls and camera-vision techniques for gesture recognition. MEDIATE was evaluated with over 90 participants with low functioning...
autism, aged between 6-12 years and with non-verbal modes of communication. The evaluation was done in four European cities: Barcelona, London, Portsmouth, and Hilversum. The study concluded that 80% of the participants achieved a sense of agency and control during the interventions. MEDIATE allowed many children to be completely on their own for the first time and feel safe in making their own choices, enjoying their behaviors, and getting recognition from the environment as to what they were doing. However, the collaborative aspects of interaction are not mentioned (Parés, Soler, et al., 2005). The work with Balloons presented in this dissertation also focuses on children with medium-low functioning autism but with an emphasis on social collaboration [Publication I].

Lands of Fog addresses issues in social fragmentation of individuals with autism by designing a game that encourages initiating social interaction using full-body interaction (Mora-Guiard et al., 2017). A virtual world with three layers that are covered in fog is projected on the floor. Participants walk on the projection with butterfly nets and work with a peer to explore, discover, and collect virtual objects by clearing away the fog. Lands of Fog uses a peephole design strategy to invoke exploration. If participants interact at an individual level, certain virtual objects, like fireflies, are revealed. However, when participants collaborate, by coming close to each other in the interaction space, larger virtual objects are revealed. In this way, Lands of Fog promotes social initiation through joint attention and shared interaction. This is one of the first examples of an application encouraging co-located social interaction through verbal or non-verbal collaboration between a child with autism and a typically developed individual (Mora-Guiard et al., 2017). Studies with Lands of Fog were conducted at two locations: a university and an inclusive school. The studies showed that results achieved in a university laboratory setting are similar to those achieved in a special school.

The Balloons application is also the first of its kind to foster social initiation through shared attention on a common object between individuals with autism and a typically developed peer [Publication I]. Furthermore, Balloons is a first of its kind in an underserved environment. It can be argued that the sophistication of Lands of Fog, visually and via tangible and full-body interaction, was made possible by being targeted towards individuals with high-functioning autism, while the simplicity of visuals and interaction in Balloons focused on children with medium-low functioning autism. Furthermore, in Lands of Fog, children with high functioning autism were paired with typically developed peers with whom they were unfamiliar. However, for studies with Balloons, children with medium-low functioning autism partnered with a typically developed individual who was familiar to them. This was intentionally done to assert that the challenges observed during the evaluations are a
consequence of the application and not of an unfamiliar presence [Publication I].

In the ECHOES project (2008 – 2012), a consortium of research labs and universities in the UK developed multimodal technology for social communication skills and focused on joint attention, turn taking, and initiating collaboration between children with Asperger’s syndrome and typically developed peers (Bernardini, Porayska-Pomsta, & Smith, 2014; Farrow & Lemon, 2011). The technology was developed using participatory design methods involving parents, children, therapists, and special educators. The technology was evaluated in four schools for eight weeks within the UK with participants between the ages of 5-7 years. The contribution of the work formed the basis of designing technology to support joint attention. The results of the ECHOES indicate measurable success for technology mediated joint attention, although they used touch-based interactions. Moreover, the game focuses on interactive virtual elements as a mechanism to promote joint attention skills, even though the peer/adult does not specifically interact with the application. Other applications include SensoryPaint, which is a multimodal application that incorporates tangible interaction and whole-body interaction for therapeutic interventions to encourage social interaction (Ringland et al., 2014).

Overall, there is a lot of focus on promoting social interaction and collaboration, including joint attentions, for individuals with developmental disabilities using technology. This can be achieved through tangible interfaces, e.g. the fish net in Lands of Fog (Mora-Guiard et al., 2017), virtual objects such as those in Pico’s Adventure (Malinverni et al., 2017), and through touch (Bernardini, Porayska-Pomsta, & Smith, 2014). The benefit of one approach over another depends on several factors, including the challenges and opportunities they provide to the target user group. According to our work, social interaction using gestures are appropriate for non-verbal children with medium-low functioning autism, as designed in Balloons. Further research is required to understand how other methods of interaction, e.g. tangible or touch, are received by children with medium-low functioning autism.

2.3 STUDIES EMPLOYING GESTURE INTERACTION FOR INDIAN USERS

There are several challenges towards introducing new technology to an underserved environment including resource constraints within the environment, for example, lack of infrastructure and access to electricity, especially in rural India (Kumar et al., 2010). There is also a large digital divide between the have and the have-nots. Digital exclusion is not only a consequence of lack of access but also gender, culture, and other socio-economic challenges (Sethi et al., 2006). Moreover, technology is
considered prohibitively expensive for individual use, therefore, it is common to share resources meant for individual use, such as mobile phones, which requires innovative localized design solutions (Marsden, Jones, & Robinson, 2014). For individuals with developmental disabilities, there are additional challenges, such as low integration and inclusivity with their typically developed peers in schools and within society as a whole. Strong cultural barriers for individuals with disabilities lead to more pronounced digital exclusion even within technology-capable communities (Desai, Divan, Wertz, & Patel, 2012). For example, an economically stable and educated parent might provide a mobile phone to their typically developed child and but not to their child with autism.

Studies show that cultural, societal and socio-economic factors largely affect “the experience of autism” (Boujarwah, Nazneen, Hong, Abowd, & Arriaga, 2011). A study of teachers’ attitude towards children with disabilities in schools in Mumbai suggests that prior acquaintance with a person with a disability is a governing factor for teachers to be more positive and welcoming towards inclusive education (Parasuram, 2006). A World Bank survey in 14 developing countries, including India, indicated a “worrisome vicious cycle of low schooling attainment and subsequent poverty among people with disabilities in developing countries” (Filmer, 2008). While there are several schools for children with special needs across India, very few of them employ technology within their classrooms and there is limited research examining the role of assistive technologies. One example is Jollymate, a digital notepad for children with dyslexia that emulates a “phonetics system of teaching letter sounds and letter formation” (Khakhar & Madhvanath, 2010).

There are relatively few studies that incorporate gesture interaction for Indian users. There seems to be a keen interest in developing technology to recognize hand gestures for Indian sign languages, using the Kinect sensor (Geetha, Manjusha, Unnikrishnan, & Harikrishnan, 2013; Ghotkar, 2015). However, technology design becomes even more challenging when working with users with low levels of literacy. Current research with users with low literacy is focused on either speech-based interfaces with IVR systems or graphical and other non-textual interfaces (Medhi-Thies, 2015), while gesture-based interaction is surprisingly unexplored in this context. In my earlier work, I explored the use of body-centric gestures to overcome the literacy barriers towards health information access (Sharma et al., 2014). A health information system with two gesture-based selection techniques (pointing to a screen and touching one’s own body part) was designed, developed, and evaluated with 37 semi-literate and literate participants in Assam (Figure 4). The results indicated a clear preference for touching in the healthcare domain. The work culminated into the following design guidelines for gesture-based health information systems: design body-centric interaction to overcome literacy and technological...
proficiency barriers, address the misconceptions of system behaviors with users not familiar with technology, understand the effects of cultural constraints on interaction, and utilize interactive virtual avatars to connect with the users. This work was a collaboration with IBM Research India and Indian Institute of Technology – Guwahati (IIT-G). It served as a proof of concept for the subsequent Chetna+11 project, in which a similar system was designed, developed, and deployed across clinics in Assam by IIT-G.

Figure 4: A participant interacting with a gesture-based health information system

Gesture applications for children in India include MathMazing, Autinet, Project-Communicate, Aarya, and Kinect-o-therapy. In MathMazing, students solve an arithmetic problem by selecting one of three options. An option is selected by dragging an avatar through a maze that has three flags at different locations, one for each option (Sahasrabudhe, Shah, Thakkar, Thakkar, & Iyer, 2012). The application was evaluated with four students between the ages of 9-12 years for its usability, however, it is unclear from the publication where this evaluation took place.

The Autinet application consists of three simple gesture-based activities (A1-A3) that were evaluated with four children with autism (Agarwal et al., 2012; Sampath, Agarwal, & Indurkhya, 2013). A1 included an interface with bubbles that are burst by controlling an onscreen stick. The stick is controlled by the user’s hand motions. A2 consists of free form drawing with red paint to introduce users to gesture-based interaction. A3 is a gesture-controlled virtual peer for children with autism designed and developed to teach imitation and turn-taking. In A3, a virtual peer waves his flag in midair to control a toy train. After he stops, the child is asked to wave their hands to control the train in a similar fashion. However, in the pilot evaluations with one child with autism, researchers observed that the child was not able to relate to the activity of waving a flag without being provided an actual piece of cloth to wave in the air. They observed that the child kept pointing to the toy train in an attempt to establish shared attention with the parent present in the session (Agarwal et al., 2012). Overall, in the Autinet evaluations, the feasibility of gesture-interaction without additional support and guidance by a moderator was considered low. The authors mention that in the future they plan to focus on tablet- and phone-based technologies because of the ease of interaction with

11 http://www.embeddedinteractions.com/Chetna+.html
touch for participants with varying motor abilities. However, they note this would increase the costs significantly.

In Project-Communicate, an augmented reality application to teach the alphabet was designed through focus group discussions with experts from an urban special school and one child with autism (Hajela et al., 2013). In the application, the user sees themselves within their physical environment, captured by the Kinect cameras, with virtual interactive objects, for instance, a card that reads A - Apple and a virtual apple. Users can zoom in to and move the virtual objects within their augmented space. However, no user studies are reported.

Aarya is a virtual avatar who guides children with autism through social interactions through different animated scenarios of shopping or at school (Sreedasyam, Rao, Sachidanandan, Sampath, & Vasudevan, 2017). The children interact through hand and body gestures to carry out Aarya’s instructions. A user study with five children with autism who used the application for one month showed improvements in social interactions. Kinect-o-therapy\(^{12}\) is focused on providing exercise and therapy games using full body and gesture interaction for physical rehabilitation for individuals with motor impairments (Roy, Soni, & Dubey, 2013). Four games were developed and evaluated for their potential therapeutic benefits and usability concerns with several patients and doctors, however, results of the evaluation are not presented.

Autinet and Project Communicate are two gesture-based prototypes for children with autism in India that predate the work presented in this dissertation. However, both focused on individual interaction, making Balloons the first co-located collaborative application with an extensive user study with children with autism in India. Both projects do mention the cost-effectiveness of using a Microsoft Kinect when shared among students of an entire school as compared to tablets. Researchers have also used available Xbox games with children with autism in India at an early childhood learning center in Bangalore, and reported improvements in motor skills (precision and stability) after an evaluation with five children (Muneer, Saxena, & Karanth, 2015).

2.4 SUMMARY
There are several benefits of employing full body or embodied gestures for interaction when designing applications for individuals with developmental disabilities. First, gestures allow for social collaboration in a co-located physical space better than say a laptop or tablet, which have limited form factor and interaction space. Second, the Kinect device and full body interaction, in general, allows for interaction without objects or

\(^{12}\) http://www.kinectotherapy.in/home.html
devices that can be thrown or broken or used to harm the self or others. And lastly, gesture interaction overcomes the challenges of interaction with physical or tangible objects, where children with autism may obsess over, for instance, “turning on and off the light of a room through the wall switch” (Parés, Soler, et al., 2005).

The Kinect motion sensor was first available at the end of 2010 for the Xbox 360 gaming console with several motion-based games. Some earlier studies using the Kinect evaluated the potential of gesture-based Xbox games for children with autism (Bartoli et al., 2013). Even though the manufacturing of the Kinect sensor stopped in 2015, research with the device (both v1 and v2) continues with several papers already published in 2018 alone.\textsuperscript{13}

Much of previous work using technology for children with developmental disabilities is focused on individual interventions in which a parent or therapist watches while a child interacts with the interface. Given the social aspects of gestures in our day-to-day communication, it is surprising then that, even with gesture-based interfaces, most of the interaction is still focused on the individual and not the social context of communication. However, almost all the previous work supports the importance of social interactions, initiating interaction, communication, turn taking, and collaboration, for children with developmental disabilities.

\footnote{Google scholar returns 2,080 results when searching for the word ‘Kinect’ and 140 with ‘Kinect Autism’ since 2018, as of 22\textsuperscript{nd} April 2018.}
3 Collaborative Technology for Underserved Children

There is a growing emphasis in HCI4D to overcome the large digital divide created by low functional literacy, lack of access, and lack of culturally sensitive material in the Global South. Research is focused on domains such as agriculture, education, and micro-finance, and across geographies, including India and Africa. Of the studies published from 2009-2014 in ten HCI conferences that are popular with researchers in the field, India accounts for more than half of the papers (Dell & Kumar, 2016). South Africa follows next with one fourth the amount of papers as those based in India (Dell & Kumar, 2016). This lack of representation at an international level is caused by lack of access to relevant resources, as is often the case in emerging markets (Gitau, Plantinga, Diga, & Hutchful, 2011). Various innovative ways have been studied to overcome functional illiteracy in accessing digital information using mobile phones in rural environments (Medhi-Thies, 2015; Toyama, 2010), text-free social networks for rural farmers (Medhi-Thies, Ferreira, Gupta, O’Neill, & Cutrell, 2015), incidental learning of a second language via subtitling (Findlater, Balakrishnan, & Toyama, 2009), job search applications for domestic laborers (Medhi, Sagar, & Toyama, 2006), non-textual and semi-textual UIs for microfinance (Medhi, Gautama, & Toyama, 2009), designing video emails (Prasad & Medhi, 2008), and using audio-icons for accessing agricultural content by rural farmers (Srivastava, Rajput, & Mahajan, 2012). Most of this work aims to harmonize graphical, auditory, and textual information to design a usable interface. However, replacing or adding pictures or audio to textual information does not benefit users with varying levels of literacy (Medhi, Menon, Cutrell, & Toyama, 2010). It ends up grouping together users with different needs and abilities into one single category, which limits their individual learning potential.
Therefore, it is important to consider literacy and learning as fluid and evolving. This is especially important when working with children.

Researchers are also looking to re-evaluate and redefine what it means to design for *development*. Current design practices are unable to support sustainable outcomes, that is, the technology is not always adopted in the long term after the study has ended and the researchers have left. Several studies have identified the importance of entertainment in making HCI4D interventions successful (Bailur, 2007; Sethi et al., 2006; Smyth, Kumar, Medhi, & Toyama, 2010; Vashistha, Cutrell, Borriello, & Thies, 2015). Moreover, there can be a gap between what rural users want and what researchers think they need, as shown in a study of kiosk-styled telecenters in India where there was a high consumption of games, digital photography, and desktop publishing content rather than healthcare and government services (Sethi et al., 2006). Studies on rural Indian voice forums also showed a high demand for entertainment-oriented content like songs and music instead of development-oriented content (Bailur, 2007; Vashistha et al., 2015). Thus, it becomes imperative to design and develop applications that cater also to user wants and aspirations to lay the groundwork for positive social change (Toyama, 2018).

Research in India is mostly focused on rural adults interacting with non-textual or semi-textual UIs (Findlater et al., 2009; Medhi, 2007; Medhi et al., 2009; Medhi-Thies, 2015). Research with rural Indian children include the ‘Hole in the Wall’ studies by Mitra et al. (Dangwal et al., 2005; Dolan et al., 2013; Mitra & Rana, 2001), Kam et al.’s work on designing digital games for English language learning on mobile phones (Kam, 2006; Kam et al., 2008; Kam, Rudraraju, et al., 2007; Kumar et al., 2010), and several speech-based English language learning applications (Larson et al., 2013; Mitra et al., 2003). Parents from low socio-economic background increasingly aspire for their children to develop computer skills, which are understood to be beneficial for employability (Pal et al., 2007). Learning English is understood also to be a high educational priority since the language barrier deepens the technology barrier as almost the entire web content in India is in English (Kam, Ramachandran, et al., 2007). This previous work with children from developing regions builds a case for computer literacy and English skills. However, inclusive collaboration, where children from different socio-economic and cultural backgrounds are raised to the same learning platform, is not well studied. In this section, I first build a case for collaborative learning and its benefits. Then I present previous research with children and technology in India with a focus on cases that promote collaborative learning.
3.1 CROSS-CULTURAL COLLABORATIVE LEARNING

The ability to work effectively and creatively with team members and classmates regardless of differences in culture and style is an essential 21st century life skill (Trilling & Fadel, 2009). Therefore, it is important for children in school to be exposed to cross-cultural collaborations to be work ready and prepared to participate as global citizens. Moreover, social media and the internet creates a hotbed of opportunities towards intercultural communication for foreign language pedagogy (Pihkala-Posti, 2012). Previous research has established the benefits of collaborative learning which also extends to online interactions. If these collaborations are multicultural, they expand the child’s view of the world and also “enhances their self-identification and empathy with different cultures” (Wang, 2011). With broadband internet connections and 4G networks becoming ubiquitous, even in resource constraints environments, it is now possible to set up synchronous cross-cultural online collaborations for children across the world, especially within closer time zones. For instance, between India and Finland, there is a 2.5 hour gap from end of March to end of October, and 3.5 hours otherwise.

Several studies on cross-cultural online collaborations have already shown that culture subtly guides interaction and communication (Vatrapu & Suthers, 2007). The usage of collaborative learning environments is dictated by the participant’s cultural model, which can be defined based on Hofstede’s dimensions of national cultures (Hofstede, 2011), namely power-distance, individualism (vs collectivism), masculinity (vs femininity), uncertainty avoidance, short-term (vs long-term) focus and indulgence (vs restraint). Vatrapu and Suthers (2007) describe the behavioral impact of the first four dimensions to online learning, and of note are the social attributes of large power-distance and collectivist societies. However, as eloquently stated by Irani, Vertesi, Dourish, Philip, & Grinter (2010), these dimensions assume culture to be static and acquired and have little to say about the “norm-shifting of technologies, social movements, or even everyday reconfigurations of practices around technology, media artifacts or experience” (ibid). From the perspective of learning within the classroom environment, a cultural snapshot of these dimensions provides an adequate starting point.

Cultural and social norms inherently influence communication and need to be addressed. In societies with a large power-distance (i.e., the expectation and acceptance of unequal social status or power distribution), communication is usually initiated by an authority figure who cannot be contradicted by virtue of respect within and outside the classroom environment. Additionally, in collectivist societies, which are focused around group life, the concept of face-saving, where the credibility and

---

14 https://opensignal.com/reports/2017/06/state-of-lte
reputation of both child and teacher is not threatened or questioned, is further strengthened. It is important to consider the consequences of face-saving in social interactions, and its effects on motivation towards learning. For example, DiSalvo, Guzdial, Bruckman, & McKlin (2014) observed that young African American adults used face-saving methods to navigate around their parents’ or peers’ expected attitude towards learning. This is because the socially accepted attitude towards technical education – to not want to be computer geeks – was contrary to their own preferences. Of note is that face-saving appears to be a universal social phenomenon. For example, Juvonen (2000) observed that both Finnish and American adolescents preferred to convey socially acceptable and approved reasons for their academic failures in order to save face.

Designing cross-cultural collaborations with Indian children and a Finnish researcher requires an understanding of how power-distance and face-saving affect communication. It can be assumed that participants from individualistic cultures exhibit more direct and explicit interaction when compared with their collectivistic peers, who tend to be more reserved in their communication (Kim & Bonk, 2006). Similar cultural differences ought to prevail between Indian and Finnish societies: the Indian culture can be considered more collectivist than Finnish, while Finland has a smaller power-distance. Therefore, the goal is to improve communication and interaction among culturally diverse participants, which in turn can potentially improve conversational language learning and collaboration (Sundqvist & Wikström, 2015). Socio-cultural barriers to communication have previously been observed in usability testing of online services with Indian adults. Next, I describe the Bollywood Method as a potential solution to overcome communication barriers for online learning and cross-cultural collaborations for children.

3.2 The Bollywood Method
The Bollywood Method was devised by Chavan (Chavan, 2005; Chavan, Gorney, Prabhu, & Arora, 2009) to overcome socio-cultural barriers towards communication during usability tests with Asian participants. The participants would not share the issues they faced or negative experiences with the interface or task, in order to be polite. This can also be attributed to the cultural concepts of face-saving and power-distance. Unable to uncover any usability issues, Chavan decided to root the tasks in dramatic stories, inspired by the Northern Indian film industry, called Bollywood. Bollywood movies usually have larger-than-life fantasy scenarios. In the dramatized scenario, participants would play a specific role that seemingly also enabled them to be more direct in their communication to achieve the task. In one example, for testing an airline’s online ticket booking system, participants were told they had to fly to Bangalore to stop the wedding of a relative to an underground goon with
the evidence in their possession. This created a character with a sense of urgency and purpose who could then share problems with the interface whilst thinking aloud.

Several studies with non-tech-savvy or low-literate Indian adults have shown that the Bollywood Method overcomes cultural inhibitions and allows users to communicate more openly and provide critical feedback (Clemmensen, Hertzum, Hornbaek, Shi, & Yammiyavar, 2008; Medhi, 2007). The Bollywood Method has been successfully used with both rural and urban Indian communities. Online pedagogies for school children are also influenced by the same cultural constraints and by the varying levels of computer experience and access. However, most of these HCI4D interventions are not collaborative or globally inclusive. Cross-cultural collaborations are also guided by the social and cultural norms of the participants’ background (Vatrapu & Suthers, 2007). Moreover, face-saving is a uniquely universal phenomena that is also experienced within teacher-student or student-student interactions (Bond & Lee, 1978; DiSalvo et al., 2014; Juvonen, 2000). Therefore, tools like the Bollywood Method provide a means to encourage communication among participants.

Within the context of culture, the Bollywood Method seems to address the challenges imposed by face-saving and power-distance. Therefore, focusing only on these two dimensions potentially overcomes the challenges towards social communication, as evident from Chavan’s work (Chavan, 2005; Chavan, Gorney, Prabhu, & Arora, 2009). However, when technology access (and, consequently, access to global culture and media) increases within an individual’s environment, it influences their social group and culture (Irani et al., 2010). This dynamic experience of culture, manifested through technology and media, requires further research. For the purpose of the research presented in this dissertation, Hofstede’s dimensions (Hofstede, 2011) allowed studying a static snapshot of culture as a starting point.

In addition to the cultural challenges towards collaboration and communication, one needs to be mindful of the differences in technology experience and fluency in the common language (in this case, English). In the work with children from Deepalaya, the Bollywood Method also seemed to assist in overcoming the lack of computer experience by focusing the child’s attention towards communication and collaboration with a sense of urgency and a tacit learning of how to use a computer.

### 3.3 Children, Computers, and Collaboration in India

The most popular work in connecting Indian children to the internet is the TED 2013\(^\text{15}\) prize-winning ‘Hole in the Wall’ project (Dangwal et al., 2005;  

\(^{15}\)https://www.ted.com/speakers/sugata_mitra
Mitra & Rana, 2001), in which children from rural and urban schools were observed while they learned to browse the internet and play games with minimal supervision and instructions in a matter of weeks. This was achieved by champion learners, that is, a child who has learned, through exploration, a new skill in a given day or week and is then sought after by their peers who want to learn the same skill. A ‘Hole in the Wall’ system consists of a screen, a keyboard with a mouse pad, and a computer boxed together as shown in Figure 5. The system has been deployed across several schools and community centers across India since 2013, and at Deepalaya in 2016 (Figure 5).

The most important outcome of the project is Self-Organized Learning Environments (SOLE) whereby, given access to technology, children working in groups discover how to use a system themselves (Dolan et al., 2013). Mitra’s work shows that children, even those from underprivileged regions, are motivated to learn how to use a computer and the internet. Furthermore, parents from low socio-economic backgrounds aspire for computer and English literacy for their children (Pal et al., 2007; Shukla, 1996). This was also observed at the Deepalaya, where parents are willing to pay additional tuition fees for computer lessons for their child (approximately €1.7/US$2 per month).

The ‘Hole in the Wall’ promotes co-located peer assisted learning and collaborations, but not online or cross-cultural collaborations. Based on the success of SOLE, a ‘School in the Cloud’ was created. Children video chat over Skype with teachers, parents, and educators from all across the world and are given a Big Question to solve. They can work together in groups (in their classroom) and use computers and the internet for the task and then share the results back with the educators. A SOLE session aims to promote exploration, discovery, and peer collaboration with minimal adult intervention. The educator, teachers, or parents are often called Grannies, and therefore this project is also referred to as the ‘Granny in the Cloud’.

---

16 This term in commonly used in ICT4D to refer to people who assist others in using technology but are themselves self-taught.

17 http://thegrannycloud.org/
(Dolan et al., 2013; Mitra, 2009). The interaction and collaboration is one-to-many, with one granny interacting with a group of children. This work is the closest to my research with CityCompass and children from Deepalaya [Publication V]. The similarity lies in usage of a computer and conversing in English. However, the difference lies in the desired outcomes: Mitra aims to enhance co-located peer collaboration and letting children learn with minimal adult intervention, that is, the children do not collaborate directly with the granny. However, my aim is to establish ways to promote online cross-cultural collaborations.

Previous studies show that mobile educational gaming provides an engaging, immersive, and rewarding experience for new skill acquisitions for rural Indian children, who have limited access to computers (Kam, 2006; Kam et al., 2008; Kam, Ramachandran, et al., 2007; Kam, Rudraraju, et al., 2007). This work is primarily focused on mobile phones catering to out-of-school children who are unable to afford or attend typical schools due to financial and practical constraints; the child is also a working member of the family. However, using games developed for a Western audience fails to provide any benefits to rural Indian children (Kam, Ramachandran, et al., 2007). By studying the game principles of rural games from different regions of India, Kam et al. have successfully developed culturally relevant mobile games for out-of-school rural Indian children (Kam et al., 2008). They observed that despite the similarity of traditional Indian games, South Indian children performed a small skit (sketch) at the beginning while the North Indian children did not. This skit performance has an uncanny resemblance to the Bollywood Method, of creating a dramatized scenario before the actual gameplay. However, the mobile games that were developed were single player or required limited peer collaboration and were not globally inclusive. This is understandable as rural India is yet to develop reliable internet access that makes globally inclusive collaboration possible.

Cross-cultural collaborations, computer skills, and speaking in English are all considered important skills to be a part of the global society of the 21st century (Trilling & Fadel, 2009). Mitra’s and Kam’s work focuses on computer, mobile, and English skills for underserved children in India. However, there is little emphasis on cross-cultural online collaboration for school aged children. My work enables underprivileged children from Deepalaya to collaborate on an online task with a Finnish researcher, using English for communication. The novelty of this work lies in its approach to combine three essential skills into one task: English language, computer skills, and cross-cultural collaborations using the CityCompass application.
3.4 SUMMARY

There is a lot of emphasis on developing educational applications for children from underprivileged communities. These applications have focused on developing computer, mobile gaming, and English skills. Most applications are also focused on the individual learner and do not tap into the potential of collaborative learning. Moreover, cross-cultural collaborations, computer skills, and English language are essential life and career skills for the 21st century (Trilling & Fadel, 2009). Multicultural experiences not only assist in developing a global perspective but also improve self-identity (Wang, 2011).

My research on collaborative language learning applications for Indian children extends the work done by Mitra et al. (Dolan et al., 2013; Mitra, 2009; Mitra & Rana, 2001; Mitra et al., 2003) of providing Indian children an online presence and global perspective, and by Kam et al. (Kam et al., 2008; Kam, Ramachandran, et al., 2007; Kam, Rudraraju, et al., 2007) of designing games for Indian children by designing globally collaborative and inclusive online applications. As social communication affects interaction styles in online learning environments (Vatrapu & Suthers, 2007), I recommend using the Bollywood Method (Chavan, 2005) to improve online collaboration when working with children from different cultures, especially when working with underprivileged communities. In the Bollywood Method, tasks are rooted in dramatized scenarios and the participant plays a character who, for instance, might have a sense of urgency towards task completion. This sense of urgency reduces the socio-cultural barriers to communication.
4 Introduction to the Research Context, Methodology, and Applications

The research presented in this dissertation is situated within the context of being underserved, more specifically, being underserved by technology known to be beneficial. Therefore, underserved here refers to lack of access to technology, which itself is a consequence of several factors. This includes a lack of access to information and support towards technology adoption, a lack of relevant or socio-culturally sensitive applications, a lack of infrastructure, or financial constraints. Working with non-government organizations (NGOs) serving entire communities provides a starting point for introducing technology to those who are currently underserved. Within a community setting, the financial costs can be mitigated by using the same technology for a large group of people. NGOs can disseminate information and provide support to the communities they work with, potentially mitigating several infrastructural issues, such as access to the internet. For the purpose of my research, I worked with two NGOs in New Delhi: Tamana and Deepalaya.

At Tamana, parents and children with autism and other developmental disabilities are introduced to technology by educators, therapists, and other specialists at their schools. At Deepalaya, parents and children from low socio-economic backgrounds are introduced to technology by teachers and computer instructors who provide specialized technical training. Both Tamana and Deepalaya have a keen interest towards technology but experience a lack of socio-culturally relevant interventions that can further benefit the children and communities they serve. Therefore, my role at Tamana and Deepalaya was to bridge this gap. The research context and
its challenges are detailed next, followed by descriptions of the
aplications that are a part of this work.

4.1 The Research Context
To situate the context of the work presented in this dissertation and the
challenges it brings, an overview of Tamana and Deepalaya, the children
they serve, and my role within the context is presented next.

Tamana Schools
Tamana is a non-government organization with four centers in New Delhi,
of which I worked with two: the School of Hope and Nai Disha. The
School of Hope was established in 1985 as the Autism Center and was one
of the first of its kind in India in two main ways. First, it provided
specialized education to individuals with autism spectrum disorder.
Second, it created awareness about autism and identified it as a
developmental disability, separate to others. The School of Hope is
considered one of the pioneers in special education not just in New Delhi,
but across India. Nai Disha focuses on developing skills for self-efficacy in
individuals with developmental disabilities, including those with high-
functioning autism. The very essence of Nai Disha, to inculcate
independence, is embedded in the way the center is run. This includes an
entire floor, in a four-story building, setup as a student hostel acting as
home to several of the students, even those whose parents reside in the
city.

Tamana centers function as schools with student uniforms, teaching
schedules, lunch breaks, group work, classrooms, course books, etc. Figure
6 and Figure 7 show two classrooms from each school. High functioning
students take lessons from the National Institute of Open Schooling
(NIOS) curriculum. NIOS, setup in 1986, is a government body, under the
Indian Ministry of Human Resources, offering vocational and skill-
oriented courses in addition to formal academic courses. The words center
and special school are used interchangeably in this dissertation.

Within the context of special schools in New Delhi, the openness of the
school environment must be noted. This includes the teaching philosophy
of collaborative or community learning and also the schools’ architecture,
with its open balconies, or verandas, that provide visibility and
communicational access between the floors (as shown in Figure 6 and
Figure 7). Therefore, no floor is individual, and the building is experienced
as a whole with activities in the ground floor visible to everyone else
within the building.

18 http://www.nios.ac.in/
19 A veranda is an open courtyard that connects rooms in a building from the inside.
Autism and other developmental disabilities are severely misunderstood in India (Hajela et al., 2013); there is a delay of almost six to seven months for an autism diagnosis for children in India as compared to their Western counterparts (Daley, 2004). Parents and children experience isolation from the rest of the extended family and friends, and are prone to misinformation and misguidance by local alternative-healers promising cures (Desai et al., 2012). Moreover, the socio-economic level of a community affects how autism is experienced by the individuals, their
parents, caregivers, and social circles (Boujarwah et al., 2011). My work with Tamana provides a first critical look at technology-mediated interventions for children with developmental disabilities in an underserved environment.

**Deepalaya Schools**

Deepalaya was founded in 1979 with the aim to educate underprivileged children, especially young girls. Their motto of *enabling self-reliance* has expanded from educating children to providing vocational training to young adults and women from underprivileged backgrounds. Deepalaya is one of Delhi’s largest operational NGOs, with three schools and six learning centers across Delhi and other neighboring states. Together with various government bodies, other NGOs, and universities, Deepalaya has influenced the lives of over three hundred thousand children. Other educational initiatives include FADA, or FAther and DAughter alliance, where fathers accompany their daughters to school on Saturdays for team-building exercises. This intervention was devised to counter the high dropout rates of girls, as it is common to only educate boys when resources are low. A community library encourages reading aloud and is open to families of students. There is also an educational bus equipped with a computer and books to function as a classroom for children who do not live close to Deepalaya centers or live on the streets.

For this work, I collaborated with the Deepalaya Learning Center in Sanjay Colony (referred to as Deepalaya in this dissertation), which is located in an industrial area. This center was previously recognized as a school but lost its *school* status when the central government changed in 2014. It was then converted into a learning center and its former students were enrolled in a public school nearby. In the public school, girls have classes in the morning and boys in the evening. Therefore, at Deepalaya, the girls come for remedial and vocational classes in the afternoon and the boys in the morning. It is understood that the quality of education provided in their specific public school is not satisfactory, from both the parents’ and teachers’ perspectives.

At Deepalaya, children over the age of 12 years can enroll in a seven-month NIIT\(^\text{20}\)-certified computer course, which costs 700INR (approx. €9/US$11). NIIT is a multinational training organization that provides training and certification for various technical and managerial skills to individuals and organizations. Deepalaya is also equipped with EduComp\(^\text{21}\) smart boards (as shown in Figure 8), which provide animated and gamified digital content customized to the national school curriculum. One thing to note is the open architecture, similar to Tamana, where the classrooms open into a courtyard (Figure 8, bottom left). Through my

\(^{21}\) http://www.educomp.com/
research, I explored the challenges towards cross-cultural collaboration for children from low socio-economic urban households.

Figure 8: Classrooms in Deepalaya with a smart board (seen fully in the top right image), (bottom left) Deepalaya courtyard, and (bottom right) entry gate to the center

4.2 The Research Methodology

The work presented in this dissertation followed a user-centered design process with exploratory user studies. Of the four applications evaluated with users, I designed and developed two: Balloons and Kirana. HOPE was designed and developed by Tamana and CityCompass by my colleagues at TAUCHI. Balloons employs gestures to promote social interaction between a child with autism and a typically developed peer or adult. Kirana employs gestures to teach the life skill of buying groceries from a local store by breaking the task down to smaller subtasks. HOPE employs gestures to teach cognitive and motor skills using spatial reasoning tasks. CityCompass promotes the learning of conversational language through collaborative wayfinding tasks.

For Balloons, the design process was carried out at the Tamana School of Hope with special educators, therapists, parents, and researchers from the school. This included two researchers, including myself, first spending one week at the school to familiarize ourselves with the environment and the ongoing therapies and interventions. Next, semi-structured interviews and focus group discussions were held to identify the challenges with the current interventions and to consolidate requirements for application design based on the hands-on experience of the experts at the school. The result of this design activity was a set of application requirements, which
were used to design Balloons. A few of these requirements were later formulated as guidelines for application design [Publications I & IV].

For Kirana, the design process included a one-day gesture-introduction study at Nai Disha’s annual school fair, where individuals with developmental disabilities and their parents, caregivers, and teachers, interacted with two gesture-based prototypes. The one-day study assisted in identifying gestures for interaction based on cognitive and motor abilities of the participants and social acceptance by the community in general. For instance, gestures that are fun within a playful environment but potentially harmful outside of it, such as punching, were identified as socially unacceptable. The design study also included short focus group discussions with parents, teachers, therapists, and researchers to identify a potential intervention for students at Nai Disha: an application that breaks down the complex tasks of buying groceries from a local store.

I note here that previous work with children with developmental disabilities usually employs several participatory design methods (Benton, Johnson, Brosnan, Ashwin, & Grawemeyer, 2011; Druin, 1999, 2002; Frauenberger, Good, & Keay-Bright, 2011; Malinverni, 2016; Malinverni et al., 2017). However, there are several issues in the underserved and HCI4D context that limit the opportunities for participatory design. First, there is high power-distance between children and adults, therefore, including children as design partners is difficult (Puri, Byrne, Nhampossa, & Quraishi, 2004). Second, children have little previous exposure to technology which “may limit their ability to envision prospective designs” (Kam et al., 2006). Lastly, it was difficult to comfortably include children with medium-low functioning autism, who are non-verbal communicators, in the design sessions where there were many unfamiliar faces. Furthermore, to avoid social awkwardness or discomfort, all user studies were conducted only with researchers the children already knew. Therefore, a user-centered design approach (focus group and expert discussions and user studies) was followed.

The four applications, Balloons, Kirana, HOPE, and CityCompass, were evaluated with groups of underserved children using exploratory user studies at Tamana and Deepalaya. At Tamana, user studies with 60 children were conducted using a mixed methods approach. In the first study, Balloons was evaluated with ten medium-low functioning children with autism at the Tamana School of Hope [Publication I]. Kirana, which was designed after a user-centered study with 18 individuals with developmental disabilities [Publication II], was evaluated with another 18 individuals with developmental disabilities at Nai Disha [Publications II & IV]. A final user study was conducted at Nai Disha with 14 participants with developmental disabilities. The study participants were divided into two groups based on their SQ (Social Quotient) and IQ (Intelligence
Quotient) scores. Both groups used the Balloons and HOPE applications [Publication IV]. An overview of the participant demographics in the user studies is presented in Table 3.

At Nai Disha, psychologists calculate SQ and IQ annually, or biyearly, for children with developmental disabilities. IQ is measured using Malin’s Intelligence test (Malin, 1969), which is an Indian adaptation of Wechsler’s intelligence scale for children (Wechsler, 1949). SQ is measured using the Vineland social maturity scale, a psychometric assessment to measure an individual’s social competence using eight categories of behavior: “self-help general, self-help eating, self-help dressing, locomotion, occupation, communication, self-direction, and socialization” (Doll, 1953). For children at the School of Hope, psychologists calculate CARS - Childhood Autism Rating Scale (Schopler, Reichler, DeVellis, & Daly, 1980).

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td># Participants</td>
<td>Age</td>
<td>Gender</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>na</td>
<td>M = 6, F = 4</td>
</tr>
<tr>
<td></td>
<td>18*</td>
<td>M = 26, SD = 5.4</td>
<td>M = 14, F = 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M = 29, SD = 7</td>
<td>F = 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M = 22, SD = 4</td>
<td>M = 4, F = 1</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>IQ</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M = 46, SD = 11</td>
<td>M = 42, SD = 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>M = 41, SD = 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SQ</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M = 58, SD = 16</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>CARS</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M = 34, SD = 3</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Tamana participant demographics
(*18 participants from the design study are not included)

At Deepalaya, two exploratory user studies with 47 children were conducted to explore the challenges with online cross-cultural collaborations, using CityCompass. Children from Deepalaya connected with a Finnish researcher located in Tampere, and I traveled to Delhi to moderate the sessions at Deepalaya. From the findings of study 1, it became evident that participants faced several socio-cultural barriers towards communication, which in turn affected their interaction with the CityCompass application and overall collaboration. The Bollywood Method, in which tasks are rooted in dramatized scenarios, is known to overcome similar socio-cultural barriers to communication for Indian adults (Chavan, 2005; Chavan et al., 2009). Therefore, the Bollywood Method was incorporated in study 2 for improving cross-cultural collaborations. Participant demographics for the two studies at Deepalaya are shown in Table 4.
# Participants 25 22
Age  
\[ M = 11, SD = 2 \]  
\[ M = 12, SD = 2 \]
Gender  
\[ M = 4, F = 21 \]  
\[ M = 9, F = 13 \]
Years studying English  
\[ M = 5, SD = 3 \]  
\[ M = 4, SD = 3 \]
Years studying Computer  
\[ M = 0.6, SD = 0.7 \]  
\[ M = 0.75, SD = 1.5 \]
# First time Computer users 3 11
# Computer Access at home 2 1
# Previous Gaming Experience 11 9

Table 4: Deepalaya participant demographics

Data gathered during the user studies consisted of quantitative measurements including task times and tasks performance metrics: pre and post-tests to evaluate mathematical abilities in Kirana and automated system logs and user experience feedback in CityCompass. Qualitative data included moderator observation, behavioral responses of individuals with autism or other developmental disabilities, interviews of parents and experts at Tamana, and participant feedback on CityCompass at Deepalaya. Several audio and video recordings were also made at Deepalaya. Data and notes collected on paper were digitized into spreadsheets for analysis. The spreadsheets, along with any pictures, videos, and audio recordings, are all stored in two online repositories that are password protected. Data collected from participants at Tamana was analyzed with researchers from Tamana, IBM Research India, and at the University of Tampere, including myself, through in-person and online discussions. Data from participants at Deepalaya was analyzed by my team and myself in Tampere. SPSS was used for statistical analysis of quantitative data and thematic analysis was used with qualitative data.

The guidelines presented in this dissertation were formulated by conducting thematic analysis on data gathered by participant interviews (including the different stakeholders at Tamana, e.g., parents, caregivers, therapists, psychologists, and researchers), and notes and observations made by all the researchers during the study (from the design to the evaluation of the applications and post study analysis). Most of the guidelines are presented in Publication IV under three similar categories: guidelines for application design, research study design, and the Indian context.

Guidelines presented in section 6.1, on technology for the underserved children, were first discussed in Publication IV under the category of designing for the Indian context for individuals with developmental
disabilities (I.I-I.III). In this dissertation, these guidelines are further expanded to underserved children, including children with developmental disabilities and underprivileged children. Based on the work with children from Deepalaya [Publication V], a new guideline (I.IV) was also added to this dissertation.

Guidelines presented in section 6.2, on designing the research study, were first devised through the analysis of the user study in Publication IV, under the category of designing for the research study (R.I-R.IV). They are expanded here to designing and evaluating applications with underserved children, including those from underprivileged regions. Two new guidelines (R.V-R.VI) were added to this dissertation based on the experiences from the user study in Publication V. This includes the challenges experienced in conducting cross-cultural and collaborative research across Finland and India for underserved children.

Guidelines presented in section 6.3, on designing interactive applications, were first developed from the analysis of the Kirana user study [Publication II], which also incorporated a few of the Balloons’ application requirements and previous work by Bartoli et al. (2014). In Publication IV, the guidelines from Publication II were expanded based on the Balloons and HOPE study and several new ones were added. For this dissertation, the guidelines in section 6.3 from Publication IV have been further explained (A.I-A.VII). Three new guidelines were added (A.VIII-A.X) from the analysis and discussion of the Balloons user study from Publication I. The applications evaluated in this dissertation are presented next.

4.3 Balloons Application - Promoting Joint Attention

Balloons promotes social interaction via joint attention, or the shared experiences over a common object, which is a crucial step toward social inclusion for individuals with autism (Bernardini et al., 2014; Farrow & Lemon, 2011). Joint attention has been linked to language acquisition and social interaction in the later stages of the neurological development of a child (Charman, 2003). The process of joint attention can be achieved in several ways, “including sharing attention (e.g., through the use of alternating eye gaze), following the attention of another (e.g., following eye gaze or a point), and directing the attention of another.” (Dawson et al., 2004). Attention can be shared or directed using several verbal and non-verbal cues, including eye-gaze, pointing, and/or speech. With Balloons, I explored the possibilities of joint attention using pointing (with dwell time) for interaction with the system. This reduced the need for complex face-to-face social communication without hindering subtle eye-gaze behaviors, although the application itself did not measure an individual’s eye-gaze behavior.
Balloons has three virtual balloons of different colors (red, blue, and purple) with the objective of selecting one of the balloons jointly with a partner. If a participant selects a balloon individually, a star is rewarded. As shown in Figure 9 (left), a golden star appears in place of the purple balloon that was successfully selected. If two people select the same balloon within a three-second dwell time, the reward is a rainbow that grows in size and has pleasant background music (Figure 9, right). In this way, selecting a balloon together has a higher reward than individual selection. Moreover, participants can decide which balloon to select in many ways: verbally by describing the location or color of the balloon, through gestures by pointing at the balloon using the onscreen hand cursors, or a combination of both verbal and non-verbal cues. This makes the application suitable for individuals with medium-low functioning autism who communicate non-verbally and encourages them towards social interaction.

Balloons was designed using a user-centered design approach with 23 people from the Tamana School of Hope, including special educators, speech and occupational therapists, researchers, parents who accompanied their child to school, and two children with high-functioning autism. Tamana School of Hope wanted to focus on joint attention intervention to address a gap within their own technology-mediated teaching. Moreover, interventions for joint attention usually include a moderator whose sole purpose is to determine where the participant is looking and whether they engage with a partner over a common object. By using virtual objects on a screen to mediate the interaction between participants and reduce the need for face-to-face communication, Balloons creates a comfortable environment for individuals with medium-low functioning autism.

4.4 Kirana Application - Supporting Life-Skills

The Kirana application breaks down the task of buying groceries from a local store in Delhi, similar to one shown in Figure 10 (left). In these local mom-and-pop stores, items are displayed on wall-to-wall shelves all
around the store and buyers usually stand outside the store and point to the items they want. The items are then taken from the shelves by the shopkeeper and put on the table, which divides the buyer from the shopkeeper. A handwritten bill is provided, based on the MRP (maximum retail price) of the items, and money is exchanged. This process is very different from a supermarket, where there are aisles of products and buyers pick and carry their items to the cashier. Several students at Nai Disha visit the local kirana store to practice buying items, however, educators find it difficult to control the real-world environment to assist students who may require a more detailed explanation of the process. The shopkeepers are also usually in a hurry to service other customers. Therefore, it was decided that an application that can breakdown the task involved in buying items from a kirana store is beneficial for students to practice with before they went to an actual store.

![Kirana store in India](image1)

![Kirana application interface](image2)

Kirana breaks down the process of buying items from a local store into subtasks: selecting an item to buy (decision making), asking for it from the shopkeeper (social interaction), looking up the price of the item, knowing if the item can be bought with the available cash (arithmetic), handing over the cash (social interaction), calculating the balance (arithmetic), and taking the balance and item (social interaction). The concept of MRP was also incorporated into the application to make it contextually and culturally relevant. Kirana’s interface simulates a typical store layout.
(Figure 10, right), where customers stand outside the store and point to items on shelves. The right side of the interface shows the money that is available for the participant to spend at the store and the left side shows an itemized bill. Participants select items on the shelves by pointing at them with a dwell time of 1.5 seconds using their left hand. When an item is selected, it animatedly moves to the tabletop counter and is added to the bill. When there is at least one item to pay for on the tabletop counter, participants can select money by using their right hand. This two-handed selection mechanism was added to encourage bodily movement and isolate item selection from money selection. The money also moves to the table upon selection and items move into the blue bag, the bill amount is reduced and, if there is a balance, it is returned to the right side of the screen. Item and money animations are sequential and easy to follow and were added to reduce the need for explicit drag and drop gestures, which were deemed difficult [Publication II]. For a detailed description of Kirana, please refer to Publication III.

Each item and money is accompanied by its spoken name and each transaction process has audio feedback in English. For example, when the balance is returned, a female voice says, “here is your balance.” The application also caters to various endings: running out of money or not having enough money left to buy an item. For these scenarios, the female voice informs you, for example, “you have no money left.” This is followed by a textual well done and an audio feedback to indicate the end of the session.

Kirana was designed using a user-centered process, in which gesture-based interaction was introduced to various stakeholders at Nai Disha. The aim was to identify a set gesture interaction vocabulary for three main interaction goals: navigation, selection, and object manipulation. The study was conducted with 18 individuals with developmental disabilities and their parents at the school’s annual fair in October 2013. It provided an informal event for researchers to interact with various stakeholders: school staff, educators, students, and parents. With an open and relaxed atmosphere, researchers had the opportunity to gain insights into the stakeholders’ expectation and identify challenges towards acceptance of technology [Publication II].

4.5 HOPE APPLICATION - COGNITIVE AND MOTOR SKILLS TRAINING

HOPE (Have Only Positive Experiences) was designed and developed by Tamana. HOPE focuses on cognitive, motor, and social skills using two learning paradigms: matching shapes, colors, and objects, and the concept of alphabet and numbers. The application provides multiple mechanisms for interaction: mouse, touch, and gestures. For the purpose of this work, matching tasks were used with the grab-drag-drop gesture for interaction.
The complexity of the gesture for interaction enables training of fine motor skills as it requires specific hand and palm movements.

The matching tasks are divided into three categories: colors, shapes, and objects. Each category consists of up to 20 screens with increasing levels of difficulty. For example, the first screen for matching colors has two green triangles and one red triangle (Figure 11, left). From screen 6 onwards, the level of difficulty increases slightly: there are two triangles (one green and one red) and one green circle, and the green triangle and green circle have to be matched together (Figure 11, center). From screen 11 onwards, another level of difficulty is introduced as all the items are of different shapes. Figure 11 (right) has a black star, a black heart, and a yellow circle, for which the correct match is the star and heart as they are of the same color.

![Figure 11: Matching Colors (left) screen 1, (center) screen 6, and (right) screen 11](Adapted from Publication IV)

The grab-drag-drop gesture consists of the following steps for matching the color of items shown in Figure 12:

1. **Grab**: select one of the blue squares by grabbing, that is, start with an open palm (shown in Figure 12, left). Move one’s hand on top of one of the squares, and then close the palm (shown in Figure 12, right). The blue square is then attached to the onscreen hand cursor.

2. **Drag**: With the palm closed, move one’s hand to the other blue square. The onscreen hand cursor follows the movement path, dragging the selected square with it.

3. **Drop**: Once both squares are almost on top of each other, open one’s palm.

![Figure 12: On screen hand cursor in (left) open palm and (right) closed palm for grab](Adapted from Publication IV)
4.6 CityCompass Application - Collaborative Language Learning

CityCompass is a virtual language learning application that uses 360 panoramic views of an actual city for collaborative tasks. CityCompass supports three interaction paradigms: mouse, full body gestures using the Kinect sensor, and virtual reality with head mounted displays. For the work presented in this dissertation, the traditional mouse interaction was used with children at Deepalaya.

In the application, participants take on the role of either a tourist or a guide, and together they navigate through a series of panoramas to reach a pre-assigned destination. Both the tourist and the guide interact freely within their views, which also contain informative clues, or hotspots, on the objects in the panoramas, as shown in Figure 13 (right). The tourist sees multiple green arrows in each panorama, where only one green arrow takes them closer to the destination (Figure 13, left). This information is known to the guide via a blue line in their view of the application, as shown in Figure 13 (right). Other green arrows either take them back to the previous panorama or to a dead-end. In a dead-end scenario, the tourist sees a culturally relevant image of the city, which they describe to the guide. The guide has to select the same image from four options. In this way, both the tourist and guide are constantly conversing with each other to complete the task. The educational and language learning potential of CityCompass is established based on previous work with BerlinKompass (Pihkala-Posti, 2012 & 2014; Pihkala-Posti et al., 2013 & 2014; Kallioniemi, Heimonen, et al., 2015; Kallioniemi, Pihkala-Posti, et al., 2015). For the work presented in this dissertation, the focus was on enabling cross-cultural collaboration for children from underprivileged regions.
5 Summary of the Publications

The work in this dissertation comprises five publications with two groups of underserved children: children with developmental disabilities and children from underprivileged regions. A total of 107 children from Tamana and Deepalaya are a part of this work. The user studies presented in the publications focus on the following research question:

*How to design and introduce collaborative educational applications for social, life, and career skills to children who are currently underserved by technology due to lack of informational, socio-cultural, or economical resources?*

This question is further motivated by the desire to explore collaborative and inclusive educational technology for underserved children. That is, applications that cater to larger group of children with different skills and abilities within the underserved context. Therefore, the research question is addressed by focusing on the following sub-goals:

- Promoting social collaboration between children with medium-low functioning autism and a typically developed peer – Publication I.
- Supporting the practicing of life-skills for children with developmental disabilities – Publications II and III.
- Furthering inclusive education to unite children with different abilities and encourage collaboration within a school – Publication IV.
- Encouraging cross-cultural collaborations between underprivileged children and a developed world peer – Publication V.

The five publications are summarized next.
5.1 Promoting Joint Attention for Children with Autism

Reference

Objective and Method
The goal of this study was to establish the potential of computer supported collaborative interventions for social interaction between children with autism and a typically developed peer or adult. Previous work with children with autism is often focused on individual interaction and online collaboration. However, co-located social interaction with a typically developed peer or adult using technology is not well studied (Mora-Guiard et al., 2016). Furthermore, interventions for joint attention (the shared attention between two people on the same object) are usually non-technical and involve face-to-face communication, which can be socially awkward for children with autism.

By using a user-centered design approach, I developed Balloons (section 4.2). Balloons focused on promoting joint attention using virtual balloons and pointing with dwell time for interaction. Two participants select a balloon together within the dwell time of three seconds using verbal or non-verbal cues, such as pointing to a specific virtual balloon or mentioning its color. In this way, participants experience shared attention through the virtual interface, with a reduced need for face-to-face interaction. Joint attention behaviors typically also involve alternating gaze between participants, however, this was not explicitly discouraged, encouraged, or studied.

A three-phase user study with ten medium-low functioning children with autism was conducted. To study whether the learning from Balloons was carried forward to real-world scenarios, phases I and III consisted of the physical balloon test. In the physical balloons test, actual balloons were taped to a wall in the school. A typically developed peer was asked to point to a balloon for the child with autism to select by physically touching the balloon, as shown in Figure 14. In phase II, participants used the Balloons application, collaborating with a researcher from Tamana, for multiple sessions within a three week period. All sessions in phase II were moderated by a researcher from Tamana to reduce any discomfort by an unfamiliar presence.
Results and Discussion

The results of phase I and III showed improvements in the time taken for a participant to touch the balloon. There was also an improvement, for several participants, in the number of virtual balloons selected jointly across sessions in phase II. Overall, the measurable results indicated a potential for the application to initiate joint attention skills in children with autism. Participant behavior and interaction with the application was observed to be similar to that observed in previous work with medium-low functioning children with autism (Garzotto, Valoriani, et al., 2014). Therefore, this leads us to believe that challenges towards social and cognitive skill development for medium-low functioning children with autism are quite universal. Thus, interventions developed in Milan could be applied to a special school in New Delhi.

![Image](image.png)

**Figure 14:** (top left and right) Physical balloon test and (bottom left and right) participants using Balloons

The implications of culture and society were experienced through different stakeholders such as parents, special educators, and therapists who were involved in the work. For instance, special educators and researchers at Tamana were motivated to use a simple application like Balloons to teach a multitude of concepts and skills, from colors and numbers to postural stability (by progressively decreasing the size of the balloons). This attitude of problem solving with limited resources is expected within the developing world or underserved context, and in India is often referred to as *jugaad* (Williams & Irani, 2010). This spirit of *jugaad* coupled with the user-centered design process, which included several of the stakeholders in the school, reduced the challenges towards technology acceptance within the school environment. Furthermore, it empowered the special educators and researchers at Tamana as social and technological agents of change.
5.2Gesture Interaction for Individuals with Developmental Disabilities

Reference

Objective and Method
Interactive technology provides several benefits to individuals with developmental disabilities, including a safe, predictable, and controlled environment. Previous research has focused on applications that promote social, cognitive, and motor skills using various interaction mechanisms such as touch, gestures, and gaze. With gestures for interaction with virtual objects, it is possible to simulate real-world scenarios with complex tasks involving social and cognitive skills. However, there is considerably little focus on applications that support life skills for real-world scenarios to foster self-efficacy. Therefore, the goal of this study was to design, develop, and evaluate an application to teach a real-life skill to promote self-efficacy in children with developmental disabilities at Nai Disha.

The work started with a user-centered design study in which two gesture-based applications were setup at Nai Disha’s annual school fair in October 2013. The applications were designed to study and formulate a potential gesture vocabulary for navigation, selection, and object manipulation. The annual school fair provided an opportunity for us to introduce gesture interaction to the various stakeholders at the school in an informal environment, where parents and their children could explore the two applications for as long as they wished. A total of 18 students with development disabilities participated in the study and feedback from 14 of their parents was collected.

The main outcomes of the user-centered design study were the development of gesture vocabulary and potential application topics. For the gesture vocabulary, it was observed that navigation was well-supported by the onscreen hand cursor and pointing with a dwell time was suitable for selection. However, object manipulations were observed to be difficult using grab-drag-drop. Furthermore, it was noted that only gestures that are socially appropriate should be used for interaction. Even though jumping, kicking, or punching are fun for interaction, they might be harmful if used outside of the play session. For the application topic, we held a short focus group discussion with teachers and parents and decided to implement a virtual Indian grocery store, called Kirana (section
Kirana breaks down the task of buying items into social interaction (interacting with a shop-keeper), decision making (deciding which items to buy), and mathematical skills (paying for the items and taking back the balance).

Kirana was evaluated with 18 individuals with developmental disabilities between the ages of 16-39 years. This large chronological age difference is common in special schools in India for several reasons: delays in getting a proper diagnosis, social taboos around special children that render them housebound, and lack of access to special schools. Furthermore, Tamana groups individuals based on their skills and abilities, not chronological ages, and is open to individuals of all ages. All the participants were unable to shop independently before their recruitment for Kirana evaluations.

The evaluation consisted of four phases. Phase I and III had pre and post mathematical tests with simple arithmetic, including single-digit subtractions and multiplications. A mathematical score, from one to ten, was assigned to each participant in phase I and III based on their performance on the test. In phase II, participants used the Kirana application for multiple sessions over a three-week period. Each session had the task of buying items for breakfast within a 100 Rupees budget (approx. £1.7/US$1.5). In phase IV, participants were taken to an actual Kirana store near their school and a researcher from Tamana observed them on three skills: decision making, social interaction, and mathematical ability. The same researcher moderated phase II and evaluated their mathematical ability in phase I and II.

Results and Discussion
The sessions with Kirana were met with excitement for playing a game. There was an average improvement of 8.3% between the mathematical scores in phase I and III. Four participants showed a decrease in performance, which is usually observed when working with children with special needs. Overall though, we believe that the application was helpful to most participants. In phase II, as participant familiarity with the application grew, their time for item selection shortened. The main goal of the study was to confirm that the learning from the application within the classroom would subsequently be transferred into a real world environment. In phase IV, which included a visit to an actual Kirana store near the school, participants were observed to be able to translate their learning from the application. Participants who felt shy would observe and repeat their more outgoing peers' interactions with the shopkeeper. They also helped each other out; for instance, one participant suggested everyone should take a bill from the shopkeeper so that an adult can verify the purchase later.
In conclusion, this study shows that applications with gesture interaction can be used to simulate and teach real-world scenarios in a safe and controlled environment. By breaking down complex tasks, participants can practice, with assistance from a moderator, before going out in the real world. Furthermore, one application can cater to several participants and for several different topics. For instance, the tasks for Kirana can include buying items for a healthy and balanced breakfast, which could initiate a discussion on the food pyramid, or buying items with a smaller or larger budget, where the focus is primarily on the concept of money. The application elements are also configurable in a text file, including all the items, their prices, and amount of money available. This flexibility in task design provides control to the teachers to personalize learning. We believe this further enhances technology acceptance and adoption in resource-constrained environments, as one application and its setup (computer and Kinect) can cater to an entire school.

5.3 Kirana: A Gesture-based Market App for Life Skills

Reference

Summary
This paper presents the Kirana application in detail and describes its interaction and system logic. It is added to this dissertation to support Publications II and IV. Kirana was designed to teach the life skill of buying items from a local Indian grocery store by breaking down social, mathematical, and decision-making skills into smaller subtasks. It employs pointing with dwell time for selection of items and eliminates the need of complex gestures by animating item movements between the different subtasks. The application is also presented in section 4.4.
5.4 GUIDELINES FOR DESIGNING GESTURE-BASED APPLICATIONS

Reference

Objective and Method
Previous research has firmly established the potential of gesture-based applications for individuals with developmental disabilities. However, very little of this research has been conducted on the Indian subcontinent, especially in resource-constrained communities. This publication presents user studies with individuals with development disabilities from Tamana, using three gesture-based applications: Kirana, Balloons, and HOPE.

Kirana, designed to teach the life skill of buying items from a local Indian grocery store by breaking it down to social, mathematical, and decision-making subtasks. It employs pointing with dwell time for selection of items and eliminates the need of object manipulation gestures by animating item movements between the different subtasks.

Balloons, designed to promote social interaction via shared experiences. It employs pointing with dwell time for selection, and two users interact with the application simultaneously.

HOPE, aims to improve motor coordination and social and cognitive skills through a series of spatial reasoning tasks, employing a gesture sequence of grab-drag-drop. In this study, the tasks are focused on matching colors, building on the concepts in Balloons, while introducing a complex sequence of gestures for selection and object manipulation.

The study with Kirana was first presented in Publication II. The study with Balloons and HOPE built upon the Kirana study with a different group of participants. The user study with Balloons and HOPE consisted of two groups of participants (as described in 4.1): one in which the participants were considered high functioning and the other with participants who were considered medium-low functioning based on their Intellectual Quotient (Malin, 1969), Social Quotient (Doll, 1953), and mode of communication (verbal or non-verbal).

The aim of the work was to explore the challenges towards learning the gesture for interaction and potential of designing an application to cater to a large group of individuals such as an entire special school. This requires applications that provide increasing levels of difficulty for the task, which
in turn can require complex gesture interaction sequences. Therefore, evaluations were conducted first with Balloons and then HOPE. The study was conducted over a period of twenty days, which included two sessions with Balloons and up to three sessions with HOPE. Tasks in HOPE were difficult at two levels – the content of tasks and the gesture for interaction. The tasks required participants to correctly match objects based on color or shape. The gesture for matching was grab-drag-drop, which required opening and closing of the palms and moving the hand across the screen. We knew the gesture sequence will be difficult for the participants and therefore the moderator, a researcher from Tamana, provided several levels of assistance: from verbal prompts to physical support. The expected outcomes were a set of guidelines for designing, evaluating, and introducing and deploying a gesture-based application for underserved individuals with developmental disabilities.

Results and Discussion
As expected, for participants in the medium-low functioning group, learning the grab-drag-drop gesture required more sessions than the participants in the other group. Moreover, the content of the matching tasks in HOPE was also more difficult than selection of a balloon in Balloons. To support the participants in the former group, the moderator provided verbal and physical assistance for the gesture and the application task. As hypothesized earlier, when designing an application to cater to a larger group of individuals, it is important to balance the gesture and task complexity. Furthermore, we observed that participants who did find the task and/or gestures difficult in their first trial were still motivated to interact, as no participant dropped out of the study.

The results of the study from Kirana, Balloons, and HOPE were compiled into guidelines. Some of the guidelines were mentioned previously in Publications I and II as observable benefits of the user studies presented in those publications. Publication IV refined those guidelines, added several more, and created three subcategories: guidelines for (a) designing a gesture-based application for individuals with developmental disabilities, (b) designing the research study with special schools, and (c) designing and introducing technology for the underserved or Indian context.

In this dissertation, the guidelines for designing interactive applications are presented in section 6.3 and are the same as the ones presented in Publication IV. Guidelines for designing and introducing technology to the Indian context (section 6.1) and conducting research (section 6.2) are refined from Publication IV and extended to underserved children, including children with developmental disabilities and children from low socio-economic urban regions.
5.5 The Bollywood Method for Cross-Cultural Collaboration

Reference

Objective and Method
Previous work with children from underprivileged regions, or resource-constrained environments, is primarily focused on mobile educational gaming or connecting children to the internet. However, cross-cultural one-on-one collaborations are rarely studied. In this work, the focus was on enabling underprivileged children from the Deepalaya in New Delhi to collaborate on an online task with a Finnish researcher, using English for communication. The novelty of the work lies in its approach to combine three essential 21st century life skills into one task: English language, computer skills, and cross-cultural collaborations.

Two user studies were conducted to identify and address the challenges faced by children from Deepalaya in collaborating with a Finnish researcher using CityCompass. I moderated the sessions at Deepalaya with the students, who collaborated with a Finnish researcher from Tampere. The setup at Deepalaya consisted of a laptop with a wireless mouse and a 3G internet connection, as shown in Figure 15. Children who could read the CityCompass instruction screen and answer questions posed to them in English regarding their name and age participated as tourists to the city of Tampere, and they had to find their way to the local street market with the help of the Finnish researcher. For the children who were not comfortable reading the instructions or conversing in English, I guided them through the application in Hindi but did not collect any data. This was done to allow all willing participants to experience the application, individually or with a friend (Figure 15, left).

Figure 15: Students at Deepalaya using the CityCompass application
Data gathered included system logs with the total time spent on the task by each participant and number of times helpful information embedded in the application, called hotspots, was accessed. Participants filled in an experience questionnaire and were also interviewed at the end of the session and asked what they thought they had learned. The questionnaire was developed by Pekka Kallioniemi, Laura Pihkala-Posti, Tuuli Keskinen, and Jaakko Hakulinen. Results of the first study showed a hesitation in both interacting with the application and communicating with the Finnish researcher. The challenges were understood to be social – related to face-saving and power-distance – and technical – related to lack of access to and experience with computers. Face-saving refers to the tendency of avoiding conflicts within social communication. Power-distance is understood to be the social status or hierarchy that creates an (unequal) status quo. For instance, in India, there is a large power-distance between a teacher and a student, such that it is impolite for a student to question the teacher (Vatrapu & Suthers, 2007).

The findings from the first study with Indian students were different than those from a similar study at a school in Tampere, which was organized and observed by Kallioniemi, Pihkala-Posti, and I. The differences between Indian and Finnish students, with respect to the interaction behavior and task completion times, can be attributed to Finnish students’ previous computer experience and a relatively flat social-hierarchy in Finnish culture. However, the goal is to encourage communication, and thus enable collaboration, for both student groups.

To overcome the socio-technical challenges to collaboration observed in the first user study with Indian students, the Bollywood Method was added to the second study. In the Bollywood Method, tasks are rooted in dramatized scenarios to overcome social challenges to communication, namely, face-saving and power-distance (Chavan, 2005; Chavan et al., 2009). In the second study, participants read aloud a dramatized story in Hindi before the collaborative task. The story introduced a sense of urgency to overcome both social and technical challenges towards collaboration.

**Results and Discussion**

Results from the two studies showed that by incorporating the Bollywood Method there was (a) a reduction in the total task times, (b) more frequent usage of helpful hints that were available in the application, (c) improved communication, and (d) a shift in the self-reported learning, from learning to use a computer to communicating in English with the Finnish researcher. The sense of urgency introduced by the Bollywood scenario reduced the power-distance between the child and researcher, and thus the children took more risks with the game elements. This includes selecting the exit to the next panorama sooner and interacting with the
helpful hints more frequently. This risk-taking further reduced the need for face-saving, thereby improving communication in the collaboration. Children were more comfortable replying with a ‘no’ rather than waiting until they could provide a positive response, for instance, when replying to the researcher’s request to find an item in the panorama. This improved flow of communication also reduced the total task time. The shift in self-reported learning is purported to be the outcome of the all the above changes in game-play behavior. Therefore, the Bollywood Method assisted in overcoming the lack of previous computer and gaming experience and reduced social and cultural barriers towards communication.

Connecting children from varying socio-economics backgrounds across the world requires an understanding of the socio-cultural challenges towards collaboration. Previous work has shown that communication is influenced by socio-cultural norms, which in turn affects online learning (Vatrapu & Suthers, 2007) and collaboration. Communication is also a critical requirement to support collaboration and for conversing in a foreign language (Gass, Mackey, & Pica, 1998). The findings of this work add to the body of research on cross-cultural collaborative learning, and specifically to inclusive collaboration, in which children from underprivileged regions are connected to the globalized world. Underprivileged here implies low technology access and usage which can be due to low parental income or limited computers at school, in both developed and developing countries.

Previous work with underprivileged children includes the ‘Hole in the Wall’ (Dangwal et al., 2005; Mitra & Rana, 2001) where children are connected to the internet for explorative learning and the ‘Granny in the Cloud’, in which teachers across the world Skype into rural Indian classrooms. However, globally collaborative one-on-one interactions are not well studied. This work presents the potential of connecting children from different cultural and socio-economic backgrounds towards common learning goals. This work also draws emphasis on developing educational applications that are globally collaborative and inclusive to children from underprivileged communities. The Bollywood Method can potentially be used as a tool for improving collaboration for children across the world.
6 Guidelines for Design, Development, and Evaluation of Applications for Underserved Children

Guidelines and practical considerations for designing, developing, evaluating, and introducing interactive applications to underserved children are presented in this section. The guidelines are divided into three categories based on their context of use. The first set of guidelines address the challenges towards technology introduction, acceptance, and adoption in an underserved context, where children are from diverse socio-economic conditions and with diverse technology experience. The second set of guidelines highlight considerations towards designing and conducting a research study to evaluate interactive applications with parents and children with diverse levels of literacy. The final set of guidelines focus specifically on designing an interactive application for children with developmental disabilities including gesture-based interaction, visual content, and structure and task flow. This last set of guidelines is seemingly universal within the context of technology design for children with development disabilities (Bartoli et al., 2014). The guidelines were derived through the research studies presented in this dissertation, and most of them are a part of Publication IV. The novelty of the guidelines lies in the research context – of underserved children – and in covering technology introduction and acceptance in that context. The section concludes with a summary of the guidelines and a discussion on the work presented in this dissertation within the domain of HCI4D for underserved children.
6.1 Guidelines for the Underserved Context

The guidelines presented in this subsection were first introduced in Publication IV, within the context of designing and evaluating applications for individuals with developmental disabilities in India [Guidelines I.I – I.III]. Here, I expand on those guidelines and introduce a new guideline (I.IV) based on continuing work with underserved children in India (Sharma et al., 2017). The guidelines I.III – I.IV are also discussed in previous literature, while I.I – I.II are novel. The guidelines address challenges towards technology introduction, acceptance, and/or adoption for children with developmental disabilities and children from low socio-economic backgrounds. They are extendable to other communities and contexts where users have little previous technology experience and the eco-system is resource constraint.

I.I Design for technology acceptance

When working in resource-constrained environments, it is essential to consider the challenges within the environment and how to address them. First, economic barriers for technologies that are too costly can be overcome by designing applications that can be integrated within schools and can be personalized and customized for use by a larger group of individuals. For instance, as previously noted, special education “centers seemed financially competent to afford at least one Kinect sensor” (Hajela et al., 2013). Second, resource constraints can also be overcome through collaborations between schools, universities, and industry partners. Third, the digital divide can be reduced by spreading awareness of the benefits of technology within schools and to parents. In India, interactive technology, especially those employing gestures, is rare and novel, and thus its acceptance remains largely unexplored. My findings suggest that following a user-centered, collaborative, and stakeholder inclusive design approach can reduce the challenges towards technology acceptance. This creates a feeling of ownership and responsibility, as also corroborated by Karusala, Vishwanath, Kumar, Mangal, & Kumar (2017).

The design and development of Balloons and Kirana followed a collaborative and stakeholder inclusive approach. Including parents and educators into the design and evaluation process created a sense of ownership and an ability to use the technology independently. This therefore increased its chances of acceptance, and, consequently, adoption. “This is especially important in the Indian context, where specialized support may not be available often, and hence parents and teachers often double up as therapists for their children” (Sampath et al., 2013).

I.II Address technology fears

As researchers and designers working with technology for underserved, and often marginalized, children, it is important to be very clear, to oneself and others, on what it means to use technology. For instance, several
parents at Nai Disha were concerned about any harmful effects of the interventions, from being potentially addictive, especially for children with autism, to effects of radiation. Together with the researchers at Tamana, we explained that all technology-mediated interventions are time bound (no more than 15-30 minutes a day) and always monitored and moderated. Furthermore, using the Microsoft Kinect eliminates the need to hold any devices.

At Deepalaya, the browser version of CityCompass was used with simple mouse interactions, that is, with technology already available at the center. Before the user studies were conducted, CityCompass was introduced to the head of the center and all the teachers present on that day. Moreover, the English and computer teachers went through the Tampere route with a Finnish researcher as was planned for the user sessions. Therefore, the staff at the center knew exactly what the user study entailed and were comfortable approaching me, if they had further questions or concerns.

Overall, it is important to discuss technology apprehensions that parents, teachers, or caregivers might have. This is especially true because there is potential for misinformation and misguidance around socially repressed topics, such as autism and developmental disabilities (Desai et al., 2012) or how technology can impact girls’ education and future (Pal et al., 2007).

I.III Give control to the educators
Providing control to the educator (teachers or special educators) to customize and personalize the intervention for each child is understood to be of utmost importance when working with individuals with developmental disabilities (Bartoli et al., 2014; Kourakli et al., 2017; Mora-Guiard, 2016; Parés et al., 2004; Sampath et al., 2013). For instance, using cars, instead of balloons, if the child likes cars (Sampath et al., 2013). In Kirana, the moderator could guide the social interaction and the task to customize the learning, even on the fly. Within the application, the number, type, and price of items, and the total amount of money in the wallet was configurable. In HOPE, the educators had access to all the screens before working with the students, enabling them to understand the sequence and complexity of the task. This could help them identify the most suitable starting level for each student.

Within the HCI4D context, there is a particular need to personalize learning goals to cater to a large group of individuals, which in turn increases the technology’s cost-effectiveness and, therefore, also reduces challenges towards its acceptance and adoption (as mentioned in I.I).

I.IV Design for socio-technical aspirations
When talking about underserved communities and HCI4D in general, the first step for designers and researchers is to identify the users’ needs. Toyama (2018) urges us to think beyond user needs and also design for
user aspirations, as this “shifts the attention from problem solving to people nurturing” (ibid), which can lay the groundwork for positive social change. Moreover, current design practices are unable to support sustainable outcomes, that is, the technology is not always adopted in the long term after the study has ended and the researchers have left.

Socio-technical aspirations differ from access to infrastructure and monetary and informational resources. This is because when designing only for the current technology access and infrastructure, we misses the opportunities provided by the ambitions and desires of the people involved. These desires may or may not be a direct result of access and goals and can arise in spite of under-developed infrastructure and lack of access. Furthermore, designing for socio-technical aspiration becomes important when considering acceptance and adoption of mainstream technology for currently underserved communities, especially children. As already shown by previous work, parents’ aspirations for their children subtly guides the opportunities they will pursue such as opportunities for learning English and computer skills for employability (Pal et al., 2007; Vishwanath, Kumar, & Kumar, 2016). This work further establishes the importance of designing for socio-technical aspirations for underserved children, where the aspirations of different stakeholders (e.g. children, their caregivers and educators) should be considered. For instance, during the work with Tamana, several parents remarked on how they would like the school to incorporate technology which is at par with the developed world (Sharma et. al, 2018). Furthermore, children with developmental disabilities across the world are known to find technology engaging and appealing (Colby, 1973; Mineo, 2009).

6.2 GUIDELINES FOR THE RESEARCH STUDY DESIGN

There are several considerations that should be considered when working with children in general, and especially within the HCI4D domain. This section provides practical solutions to cover important social, ethical, and research agendas when designing the study. R.I – R.IV were first introduced in Publication IV for individuals with developmental disabilities but are extendable to designing and evaluating interactive applications with underserved children. R.V & R.VI present challenges in conducting cross-cultural and collaborative research across countries, cultures, and continents, involving both children with developmental disabilities and children from underprivileged regions. Guidelines R.IV - R.VI are also discussed in previous literature, while R.I – R.III are novel to this work.

R.I Design for diverse learners for inclusion-within

Focusing on interactive technology to nurture collaboration, instead of competition, fosters social inclusion and creates a sense of unity among
diverse groups of learners. Additionally, this *inclusivity* provides an opportunity to accept the strengths and challenges of other students. Using one application or device within a center for all groups of learners also reduces technology costs. Moreover, as suggested by Guiard, placing technology within a special school for all to use can also potentially strengthen “the social relationships between members of the school community” (Mora-Guiard et al., 2017). Mitra’s work also shows promising benefits of children’s self-organized learning environments (SOLE), where a *champion* learner of the week can potentially tutor their peers (Dangwal et al., 2005; Dolan et al., 2013; Mitra & Rana, 2001).

Fostering inclusion and collaboration becomes especially important when individuals have varying levels of reading and language skills within the same grade level, as is typically observed in an underserved environment (Kam, Ramachandran, et al., 2007). Both Tamana and Deepalaya cater to children who are not only technologically underserved but are also on the fringes of mainstream school education. Therefore, it is also highly likely that children in the same class or grade level have varying abilities and that they are grouped into classes based on their abilities rather than chronological age. For instance, as discussed in section 4.1, chronological ages of Kirana participants varied from 16 to 39 years. Similarly, user studies with Deepalaya included children who were between 9-18 years old. Therefore, all individuals were encouraged to participate in order to promote social bonding regardless of their previous technology experience.

**R.II Promote self-efficacy**

As an organization, Deepalaya is committed to enabling self-reliance among women and children from low socio-economic rural and urban communities in India. Self-reliance for children specifically implies providing education, training, and skill development that enables them to be employment ready for the information age. For Tamana, it is important to impart life skills to children with developmental disabilities to promote self-efficacy. Moreover, self-efficacy, or the feeling thereof, can further drive student’s motivation to learn (Zimmerman, 2000).

Interactive technology and applications should therefore empower the child by involving them in decision making, and thus promote self-efficacy. This can be achieved in several ways. First, by supporting globally collaborative applications like CityCompass, where underprivileged children can connect with individuals from the developed world for specific learning goals. This enables them to acquire computer, English, and cross-cultural collaboration skills; all of which are deemed as necessary 21st century life skills (Trilling & Fadel, 2009; Warschauer, 1997; Yang, Kinshuk, Yu, Chen, & Huang, 2014). Second, through collaborative applications for children with developmental disabilities that incorporate socially appropriate gestures and interaction
to ensure that the learning, when translated to a real-world scenario, is acceptable and does not further isolate the participant. This is particularly relevant in the developing world where social awareness about individuals with disabilities varies greatly. Kirana does the necessary groundwork in preparing individuals with developmental disabilities for purchasing items of need from local stores, promoting self-efficacy. Moreover, individuals with developmental disabilities, such as autism, display a tendency towards repetitive behavior and teaching socially unacceptable gestures might increase their isolation, even if the gesture is fun during gameplay (Bartoli et al., 2013, 2014).

R.III Address socio-ethical challenges
Any resolution of conflict between the research procedure and the participants’ interest should always be in favor of the participants. When working with children, especially those who are new to technology, moderators should accommodate individual needs of each participant.

With Tamana, there were several socio-ethical research challenges that were resolved with the participant’s best interests in mind. This included forgoing a control group to allow equal opportunity to all children towards participation and providing physical assistance for gestures or interaction as and when requested. In the study with HOPE [Publication IV], the initial goal was to plot a learning curve of the gesture – grab-drag-drop – for which it was important to provide limited assistance. Yet, the moderator quickly realized the need for several levels of assistance. Moreover, several participants showed an interest in using an application after the study ended, which researchers at Tamana gladly facilitated.

With Deepalaya, participants were recruited opportunistically with the help of their classroom teacher. As mentioned in R.I, children with diverse skills are present within the same classroom, thus, at the beginning of each session, participants were asked to read aloud the CityCompass instruction screen. The ones who could read were asked if they would like to interact with a Finnish researcher. The others were asked if they wanted to play, and then they ‘played’ informally with the moderator (data not included in the study presented).

R.IV Provide assistance for gestures
In interactive technology, it is well known that gestures can induce fatigue and require uncomfortable motor movements. When working with children with developmental disabilities, who have varying motor and physical strengths, assistance should be provided to reduce fatigue. In the user study with HOPE, the moderator assisted participants who felt tired by supporting their elbow and guiding hand movement. Providing physical support for gesture-based interaction was also suggested by Sampath et al. (2013), further adding the need “for the possible presence of a caregiver [to guide] the child”. As mentioned in R.III, support for
interaction should be provided regardless of the task times being measured to ensure that the child is not hurt in any way.

R.V Obtaining informed consent
For each of the user studies conducted, the sessions with children happened during their school day and as a part of their school activities. Therefore, informed consent was sought from the heads of the school and their HR departments. At Deepalaya, I presented my research topic, plan, schedule, and demonstrated the application. I then shared a consent form, asked them to review it, suggest changes, and sign it. With Tamana, a joint study agreement was signed between Tamana and IBM Research India. Sessions at Tamana were conducted only by researchers from Tamana. Furthermore, both Deepalaya and Tamana have guidelines and a vetting process for volunteering with them.

However, one must note here that children should still be explicitly told that they can stop participating any time without any reason to respect their “autonomy and the principle of voluntary participation, regardless of whether a guardian's consent has been obtained or not”22. This leads to the concept of fluidity of consent, that is, consent being an ongoing renegotiation process for each participant as they experience the experimentation process (Sin, 2005). Therefore, it is essential to make sure that the participants are comfortable during the entire session by asking about and observing their level of comfort.

R.VI Respecting participant’s privacy and the research environment
Working with vulnerable and marginalized participants requires special considerations of keeping their personal data private and confidential. When it comes to sharing videos or pictures of the participants, especially since most of the participants enjoy having their picture taken, there is one main challenge: providing a neutral image. It is important to avoid showcasing a specific Western image of an underprivileged context - for instance as slum dwellers. This is not the context or premise of my research and is a negative portrayal and a misrepresentation of the hard work and positive attitudes of students, teachers, and NGOs. However, as a researcher, it is important to remain objective and neutral, and not showcase images that are seemingly positive. All visual material should be filtered before sharing by blurring out faces, removing any identifiable information, and ensuring a neutral image of the environment is portrayed. Furthermore, all material (videos, images, papers) is also shared with the people involved in the work at the school.

6.3 Guidelines for Applications Design

The guidelines presented in this subsection are focused on designing applications that support socially collaborative interaction. Guidelines A.I – A.VII were first presented in Publication IV, and A.VIII – A.IX have been added to this dissertation. The guidelines are seemingly universal: they are separate from the context of culture and are generally more applicable to designing for children with developmental disabilities, as also supported by previous work in this domain (Bartoli et al., 2014; Malinverni, Mora-Guiard, & Parés, 2015; Sampath et al., 2013). In fact, it can be stated with some degree of confidence that interaction needs of children with medium-low functioning autism are universal. This is because the interventions focus foremost on social interaction, namely, joint attention, imitation, turn-taking, and social initiation. All but one guideline presented in this subsection (A.IX) is supported by previous work: previous work suggests designing gestures that are playful in nature (Bartoli et al., 2013), but this was frowned upon by the stakeholders at Tamana.

A.I Provide a clear start and end of gameplay

Provide a well-defined start and end of a session to allow the participants to understand when a task is over. A clear start indicates when to focus on the application and begin interaction. Likewise, a clear end signifies when to stop the interaction. This can be especially useful to streamline the sessions and define the session duration, which supports schedule charts used with children with autism to establish a routine. Furthermore, having a clear end state supports turn-taking within a socially collaborative environment. This also potentially reduces excessive attachment with technology and discourages repetitive gestures and behaviors, which are a common concern, especially for children with autism.

In Balloons and Kirana, the application is revealed only when a participant is within the area most suitable for interaction with the Kinect. When there is no participant in that area, the screen is black to indicate that the system will not respond to any inputs. Kirana also addresses several end-of-play scenarios (for example, running out of money to buy another item) by stating the reason for the end, and rewarding participation with both a visual and verbal “well done”.

In previous work, researchers have emphasized the need for exploration to promote imagination in children with autism (Mora-Guiard et al., 2016, 2017). In Lands of Fog, there was a time limit set for each session, even with open-ended play which focused on exploration and collaboration with an unknown peer. This goes well with the general understanding that children with autism require that schedules be known beforehand. Moreover, Lands of Fog was designed for children with high-functioning autism with the aim to introduce flexibility and thus the open-ended play.
To accommodate children with medium-low functioning autism, it is essential to provide a clear start and end to the session.

A.II Provide feedback with multiple modalities
Both visual and auditory feedback should be provided because participants have varying levels of visual and auditory preferences, sensitivities, and capabilities. This also supports multiple environmental contexts, for instance, if the interaction is in a noisy classroom, then visual feedback compensates for the audio issues. Furthermore, multimodal feedback also provides multiple stimuli for attention (Bartoli et al., 2014). Providing multimodal feedback also allows for diverse stimuli that makes an application suitable for a larger group of participants.

In Balloons, the reward of the rainbow had both visual (growing in size) and auditory (music) feedback. The star reward was intentionally subdued, with no auditory feedback, to highlight the rainbow reward to encourage participants to jointly select a balloon. In Kirana, when a virtual object is selected by pointing and dwelling, the object grows and its name (milk, eggs, or five rupees) is said out aloud by the system. In HOPE, the system provides visual and auditory clues for object selection. However, for the interaction gestures, the moderator provided verbal, visual, and physical prompts to assist the participants. In the future, verbal and visual prompts can be incorporated into the system.

A.III Offer rewards and positive reinforcements
Within Tamana, the educators and experts emphasized the importance of rewards and positive reinforcement, regardless of the task performance, to encourage participants to interact and to sustain their interest and engagement. This is also supported by previous studies by M4ALL (Bartoli et al., 2014; Mora-Guiard, 2016; Mora-Guiard et al., 2016, 2017; Parés et al., 2004). This can be achieved by adding auditory applause at the end of a selection or session. The system can also provide automated verbal appreciations, like saying “well done”. Since moderators are present in a session, and the interactions themselves are socially collaborative in nature, depending on the comfort level of the participant, there can be high-fives, clapping, or other physical activities that the participant likes and finds interesting. An end of the session reward can also be playing a favorite mobile game, which may be unrelated to the application.

In Balloons, an individual selection is rewarded with a star, and a joint selection, which indicates some form of social collaboration, is rewarded with a growing rainbow with playful music. In Kirana, the end of a session is rewarded with an applause and a well done message, regardless of the interaction with the application and whether any items were bought. In HOPE, a correct answer is rewarded with flying balloons and an auditory applause. Furthermore, the moderators also provided verbal
appreciation and encouraged interactions and steps taken towards a task, regardless of the end result.

A.IV Evolving task difficulty levels with easy transitions
Applications that provide varying levels of difficulty and enough repetitions allow educators to customize learning goals and cater to a large group of participants (Bartoli et al., 2014). By varying the level of task difficulty, participants can be motivated to spend more time with the application to reach the next levels. However, one must note here, participants and moderators should be allowed to choose the level they want to begin with, as this ensures that levels that are too easy are skipped. In applications that have user profiles, this is achieved more easily, but this was not the case in our studies. By allowing for repetitions within the same level of difficulty, participants are able to practice more with the same task if they find it too difficult to move forward.

In Balloons, the selection allowed for only two levels of task difficulty. However, it was designed specifically for medium-low functioning participants with autism with limited social and verbal interactions. In Kirana, a moderator could define the task with varying levels of difficulty. For instance, in the user study the aim was to buy a list of items within a given budget, and the list was usually provided. A task requiring participants to ‘buy a nutritional breakfast’ without a list of items, can also be completed with the Kirana application. In this case, the participants need to account for nutritional aspects of the items they are buying. In HOPE, the matching colors task starts with three items of the same shape: two green triangles and one red. From the sixth screen on, the difficulty increases by having two items of the same shape and one of a different shape. Then, from the eleventh screen, another level of difficulty is introduced by having all items of different shapes. In this way, the level of difficulty increases every five screens.

A.V Provide serial and structured content
Complex tasks should be broken down into a sequence of smaller and manageable sub-tasks that follow a specific structure. The transition between sub-tasks, using animations, sounds, or other audio-visual content, should also be easy to follow. This allows moderators to identify which specific sub-step requires more practice. Furthermore, designing serial and structured content also supports guidelines A.I - designing a clear start and end of gameplay and A.IV - evolving task difficulty levels with easy transitions.

In Kirana, the task of buying groceries from a local store was broken down into smaller sub-tasks of deciding what to buy (decision), checking the wallet (and current bill) to see if there is enough money to buy an item (arithmetic), selecting the items by pointing and dwelling on them (social interaction), and then paying for the items using the wallet and calculating
the balance (arithmetic). Transitions between each of the sub-tasks were participant dependent, and the tasks could be completed in various sequences, for instance, selecting and paying item-by-item or for several items at once. Moreover, the buying animations, related to movement of money from the wallet to the table, and then the return of the balance were slow and sequential [Publication III].

A.VI Simulate real world scenarios

The overarching benefit of gesture-based applications is the potential to simulate real world social interactions in a classroom, which is a safe and controlled environment. This is achieved by incorporating socially and culturally appropriate gesture interactions and by designing content to match real world activities. When compared with other modalities, gestures are inherently embodied (Malinverni et al., 2012), creating an immersive and engaging experience that can be utilized to encourage real world interactions. They also provide a comfortable personal space for socially collaborative interactions. Touch screens have a limited form factor to support simultaneous interaction of multiple participants. Therefore, gesture-based application can be designed to simulate a real world scenario through their content or interaction, or both.

In Balloons, two participants jointly select the same virtual balloon for the social experience of joint attention. In Kirana and HOPE, even though designed for a single participant, the moderator plays a crucial role in the interaction. Moreover, because the interaction is performed at a distance from the screen, the application flow and participant interaction is visible to an audience, who can prompt and encourage the participant and become involved in the activity.

A.VII Balance gesture and content complexity

Educational applications that employ novel methods of interaction inherently have two learning paradigms in one task. The first part is learning how to interact with the application – in this case using pointing or drag and drop gestures. The second is solving the task – for example, matching colors in HOPE. Therefore, when designing an application, it is essential to incrementally change the difficulty level of either the content or interaction method, but not both at the same time. This ensures that the learning is translatable between tasks, which is important to sustain interest and engagement and provide a sense of achievement (Bartoli et al., 2014).

In Balloons, the gestures and content are intentionally kept simple to encourage social interaction among the participants without being distracted by objects on the screen. In Kirana, the gestures for interaction were restricted to pointing and dwelling to compensate for the complexity of the content, which incorporated three different sub-tasks: decision making, arithmetic, and social interaction. Similarly, in HOPE, the task’s
complexity increases incrementally, while the grab-drag-drop gesture for interaction, albeit slightly difficult to begin, stays the same. In the user studies, participants were first introduced to Balloons for practicing gesture interaction before they used HOPE. Therefore, the concept of selecting a color was translatable from Balloons to HOPE, while the gesture for the selection increased in complexity, from pointing with dwell time to grab-drag-drop.

In previous research, one application in MEDIATE focused on mid-air gestures to support a virtual Kite, however, such level of motor control was found to be difficult for children with low-functioning autism, and therefore it was not used in the final design (Parés, Soler, et al., 2005). However, the researchers then explored simple behaviors, like moving towards or away from a screen for interaction, hoping the user would understand that she is control of the situation (Parés, Soler, et al., 2005).

**A.VIII Reduce visual clutter to focus attention**

The content of the application needs to be carefully designed to ensure that visual and auditory stimuli is pleasing and not distracting. For instance, visual clutter (unnecessary virtual objects or animations) should be avoided. Moreover, object movements and animations should be sequential, making it easier to focus attention on one task or object at a time. However, this does not imply all visual content should be static or bland. Include colorful graphics with smooth visual movements and illustrations to captivate attention for longer time spans (Bartoli et al., 2014). It can be argued that a busy background might arouse curiosity and appeal to participant’s personal preferences. However, it is safe to assume that a simple background makes it easier to focus on objects of interest.

In Balloons and Kirana, virtual objects are placed on a plain light-colored background, such as white or yellow. This highlights the colorful virtual objects in the application. In HOPE, the background is colored but plain.

**A.IX Design socially acceptable gestures**

During the user centered design study [Publication II], it was evident that participants enjoyed playful gestures such as a mid-air punching or jumping. However, in the discussions that followed, the parents, educators, and other experts were quick to point out the necessity to encourage only socially acceptable gestures in our applications. This ensures that the learned gestures, if used out of context, are not socially isolating or physically harmful to the participants or others around them. For instance, gestures such as kicking or jumping that are fun for gameplay can be dangerous outside of the session. Therefore, all the applications presented in this dissertation focused on upper body hand gestures and are also socially relevant, for instance, pointing to an object for shared attention.
Previous research with Xbox games for children with autism have successfully employed fun and playful gestures, such as jumping (Bartoli et al., 2013; Muneer et al., 2015). This is the only guidelines that is contradictory to findings from previous research on designing gesture-based interaction for children with developmental disabilities. However, in MEDAITE, one of the first full-body interactive environments for children with low-functioning autism, researchers expressed their desire to focus on natural gestures which “are ergonomically, culturally and socially adequate for the type of users and the type of activities” (Parés, Soler, et al., 2005).

**A.X Encourage collaboration and social interaction**

Previous research on gesture-based applications for participants with developmental disabilities is focused mainly on individual interactions (Bartoli et al., 2014; Kourakli et al., 2017; Parés et al., 2004). However, gestures allow for social interactions and shared experiences for non-verbal communicators. Moreover, social isolation and lack of motivation for social interactions are challenges that gesture-based applications inherently overcome. Therefore, as the last, but certainly not the least, guideline for application design, I urge researchers, designers, educators, and other experts working in this domain to design applications that are socially inclusive and provide several opportunities to collaboratively interact with peers, family members, and other caregivers. This can be achieved by multi-user interactions, like in the case of Balloons, or even single-user interactions that require a moderator or peer to be involved in the task, as with Kirana, or even as an audience, as with HOPE. Moreover, “to promote behaviors related with social requests, it is advisable to design cooperative game mechanics where different resources are distributed between the players to achieve a common goal” (Malinverni et al., 2017).

**6.4 Summary of Guidelines**

This dissertation provides an exhaustive list of guidelines for fellow researchers, designers, and practitioners working with technology for underserved children, on how to design, deploy, and evaluate similar applications within a school environment. Table 5 provides a summary of the guidelines, which also briefly includes their justification as explained through section 6.1-6.3.
| Technology for the underserved context | I.I Increase technology acceptance by designing for large groups of children to mitigate costs and including stakeholders in the process. I.II Address apprehensions of parents, teachers, or caregivers towards technology usage to reduce misinformation and misguidance. I.III Empower educators by giving them control over the different ways an application can be used. I.IV Design for parental aspirations towards technology for their children instead of only focusing on a limited set of needs. |
| Considerations for the research study with underserved children | R.I Design for a diverse group of learners and for inclusive learning to create a sense of unity among children. R.II Promote self-efficacy and design to empower children in their interactions with technology. R.III Resolve conflicts between research procedure and the participants’ interest or needs by favoring the participants. R.IV Provide assistance for gestures and other interaction mechanisms as and when requested or needed. R.V Even with informed consent from the head of the school, make sure the participant is comfortable during the session and can leave anytime. R.VI Respecting participants’ privacy and the research environment by remaining neutral and objective towards the work and images used. |
| Designing interactive applications for children with developmental disabilities. | A.I Provide a clear start and end of gameplay to support turn-taking, routines, and schedules, and to reduce attachment to technology. A.II Provide multimodal feedback to cater to varying levels of visual and auditory preferences, sensitivities, and capabilities. A.III Offer rewards and positive reinforcements to keep participants motivated and engaged. A.IV Evolve task difficulties with easy transitions to allow educators to customize learning goals and cater to a large group of participants. A.V Allow participants to learn complex tasks by breaking them into a sequence of smaller, structured, and manageable sub-tasks. A.VI Utilize the inherent potential of gesture-based applications to simulate real world social interactions in a safe and controlled manner. A.VII Balance complexity between the content and interaction while learning to use novel applications. A.VIII Reduce visual clutter to focus participant attention on objects of interest that support interaction. A.IX Design gestures and interactions that are socially acceptable also outside of the context of the application. A.X Encourage collaboration and social interaction among children with developmental disabilities and their peers, parents, and caregivers. |

Table 5: Summary of the guidelines for designing, developing, evaluating, and introducing collaborative applications to underserved children.
6.5 Discussion

The guidelines presented in this dissertation aim to provide practical starting points for researchers, educators, and other practitioners working within the field of educational technology for underserved children. The reasons for being underserved by technology are several. This includes the prohibitive cost of the technology, inadequate supporting infrastructure, lack of awareness due to low parental literacy levels, and rigid socio-cultural norms. For instance, parents raising a child with disabilities face several socio-cultural issues that delay or restrict the child’s opportunities towards special education (Daley, 2004; Desai et al., 2012). Underprivileged children living in informal, urban settlements attend inadequate or poorly funded public schools. The work presented in this dissertation focused on children from NGOs in New Delhi, however, the guidelines are applicable to underserved children and their families, who face similar challenges towards technology access, within and outside of India.

A large part of previous research on underserved children and technology is focused on individual interactions. Technology that is focused on the individual, although beneficial for specific skills like introducing the concept of self, can potentially isolate a child from their social group and from the globalized world. Furthermore, acquiring computer skills and cross-cultural collaborations open up many doors for employment in the future (Trilling & Fadel, 2009). Therefore, the focus of my research was on exploring the potential of collaborative applications for underserved children. The novelty of this work lies in its dual level of collaboration: the collaborative interaction using the application and the collaborative nature of the research. Next, I discuss designing collaborative educational applications for underserved children based on the experiences with children at Tamana and Deepalaya.

The case for collaboration for children at Tamana

Research studies in India on technology for children with developmental disabilities are sparse and primarily focus on tablet-based interventions (Vellanki et al., 2016). Tablets are potentially less collaborative than gesture interaction in co-located technology sharing environments. Children with developmental disabilities face several social, cognitive, and learning challenges. The potential of learning gestures or full-body interaction is strongly supported by the embodied cognition theory (Malinverni, 2016; Malinverni et al., 2012). Gesture-based interactions can be highly social and interactive and can be successfully used to mediate co-located social interactions, as in the case of Pico’s Adventure (Malinverni et al., 2017), the Pictogram Room (Herrera et al., 2012), Lands of Fog (Mora-Guiard et al., 2017), and Balloons [Publication I]. However,
Pico’s Adventure, Lands of Fog, and Pictogram Room are geared towards medium-high or high functioning children with autism. While Balloons was specifically designed to promote social interaction between a medium-low functioning individual with autism and a neuro-typical peer or caregiver. It focused on joint attention skills that were supported by both verbal and non-verbal interactions, eliminating the need for face to face communication that might be stressful. Joint attention interventions have employed various interaction mechanisms, from touch (Bernardini, Porayska-Pomsta, & Smith, 2014) to tangible (Mora-Guiard et al., 2017). Gestures eliminated the need to hold an object or tablet and created a more open space for interaction, which was preferred by Tamana.

From the evaluation experiences of Balloons, it was also evident that the educators at Tamana could potentially expand the application to other topics, for instance, teaching the concept of colors, numbers, types of objects, and even hand-eye coordination and fine-motor skills by decreasing object size. The underlying principle of promoting the social skill of joint attention would still remain. Similarly, Kirana can be expanded to teach the concepts of money and budgeting and buying items that provide a nutritional and balanced meal by changing the virtual objects and their costs. Although Kirana was designed for individual interactions with the system, the role of the co-located moderator supported collaboration. For instance, the moderator created the task for the children and provided various levels of assistance.

In HOPE, the interaction was individual, but the application was used with an audience and with the moderator providing both verbal and physical support for the task and gestures. Furthermore, with HOPE, medium-low functioning to high functioning children with autism could be encouraged to socialize and collaborate, making it inclusive. While identifying a gesture vocabulary for individuals with developmental disabilities [Publication II], the gesture-sequence of grab-drag-drop was understood to be difficult to learn, however, high functioning individuals learned it in one session while medium-low functioning individuals took several more. Ideally, the gesture should be intuitive and natural with the shortest possible learning curve, yet there should also be a gradual increase in the complexity of a task. Moreover, when designing applications that are inclusive and collaborative, individuals will display different learning behaviors and times.

At Tamana, I also encouraged audience participation and stakeholder inclusiveness with regards to parents, peers, special educators, and other therapists. This was important in the context of my work, as in India the concept of community is strong, being a generally collectivist society (Hofstede, 2011). Socially collaborative and co-operative interactions
reduce the social isolation already experienced by children with developmental disabilities.

**The case for collaboration for children at Deepalaya**

Cross-cultural collaboration, computer skills, and conversing in English are important 21st century life and career skills for the workers of tomorrow (Trilling & Fadel, 2009). Research on educational technology for underprivileged children in India is primarily focused on mobile applications for English language learning (Kam et al., 2008; Kam, Ramachandran, et al., 2007; Larson et al., 2013) and connecting children to the internet for self-organized learning with minimal teacher supervision (Dolan et al., 2013; Mitra & Rana, 2001). Although the ‘Hole in the Wall’ and the ‘School in the Cloud’ encourage co-located peer collaboration (Dolan et al., 2013; Mitra & Rana, 2001), there is little focus on online cross-collaboration, especially one-on-one. For instance, the ‘School in the Cloud’ promotes and encourages children from urban and rural India to interact and converse with an educator from another culture. However, the role of the educator is limited, and interactions are one-to-many. This work is the closest to my research with children at Deepalaya but it differs in two ways: I connect children one-on-one with a remotely located foreign peer or adult and focus on online collaborative tasks instead of only searching for answers to a problem on the internet.

Students from low-income households with limited computer experience may face socio-technical challenges towards cross-cultural online collaborations. Previously known challenges towards collaboration and communication include social (power-distance and face-saving) and technical (lack of or limited computer experience) challenges. During usability studies with Indian adults, Chavan devised the Bollywood Method to overcome the social barriers to communication, namely, power-distance and face-saving (Chavan, 2005; Chavan et al., 2009). In work with children from Deepalaya, grounding the task in a dramatic scenario reduced the observed socio-technical challenges towards collaboration.

The adaption of the Bollywood Method, as a scenario to provide a specific context and agency (and urgency) to an online collaborative task, is similar to other scenario-based approaches for promoting user participation. The method can be implemented in the application in several ways, such as animation, video, or even a simulated telephone conversation. In fact, the application could also be adapted to incorporate encouragement or guidance to increase communication. The findings from the user studies can extend to other cultures and regions, as the social barrier of a large power-distance society is not only common to Indian cultures but is prevalent in most Asian cultures (Hofstede, 2011), and face-saving is understood to be universal (Toomey & Kurogi, 1998; Vatrapu & Suthers, 2007). Moreover, the digital divide is not a binary construct between the
haves and have-nots, but rather a complex socio-economical stratification within and among both developed and developing countries (Livingstone & Helsper, 2007; Rao, 2005; Warschauer, 2003). Therefore, it stands to reason that the Bollywood Method can be adapted to benefit online environments for cross-cultural collaboration for children across the world.

Within the domain of interaction design for children, there are several scenario and narrative-driven design approaches for learning, which are similar to the Bollywood Method. For instance, research has shown that storytelling approaches can promote students understanding of “curricular content and improve their technical, collaboration, and communication skills as they engage in long-term storytelling projects” (Sadik, 2008). The Bollywood Method is closest to the concept of Forum Theater in which experiences and narratives are communicated through short plays (Muller, 2003). Previous work in rural South India has shown that children performed a short play or skit before starting a traditional game (Kam et al., 2009). The skit closely resembles our adaptation of the Bollywood Method, that is, providing a background scenario to a game-task only before starting that task, and not during.

**Overcoming the challenges of an underserved context**

In this dissertation, the term *underserved* is used extensively to refer to several known challenges that are inherent in developing regions. First, there is a lack of resources in terms of infrastructure and blanket access to technology. The introduction of technology in a resource constrained environment incurs several costs: monetary, time, and effort. Second, there can be several social-cultural challenges towards technology access, including traditional gender roles or social taboos surrounding disability. Furthermore, the relatively collectivist nature of such a society makes it difficult to overcome rigid social norms. For instance, at Tamana, one of the first interventions after a confirmed autism diagnosis is mother-child bonding. This is because children with developmental disabilities and their families are affected by the limited understanding of autism, and other neuro-developmental disabilities (Daley, 2004; Desai et al., 2012). The strong social taboos associated with raising a child with a disability also affect the *experience of autism* for the individuals, their parents, and care-givers (Boujarwah et al., 2011; Desai et al., 2012). On the flip side, researchers can capitalize on the strong sense of community by understanding stakeholders and defining their role towards the interventions. This creates a sense of ownership that improves technology acceptance and adoption in the community (Karusala et al., 2017).

Previous work with children with developmental disabilities has established the learning and therapeutic potential of technology for social, cognitive, and motor skills acquisition. Technology is appealing because it provides a safe, controlled, and predictable environment for interventions.
However, most of the work is focused primarily in the developed world, with limited work with children with developmental disabilities in the underserved context to build upon. Benefits of the work to individuals with developmental disabilities in the developed world are strongly established, dating as far back as 1973 (Colby, 1973) to the growing crop of studies in the last five years, when the Microsoft Kinect motion sensor was first introduced in 2012. From the work presented in this dissertation (section 6.3), it can be argued that technology-mediated solutions for reducing social fragmentation of individuals, especially those on the medium-low functioning spectrum of autism, seem to be universal. Therefore, the benefits of technology for children with developmental disabilities established by research in the developed world are applicable also to those in underserved regions.

At Tamana, challenges towards technology introduction and acceptance mostly arose from a lack of informational and socio-cultural access. Even when parents, school experts, therapists, and other caregivers of individuals with developmental disabilities were somewhat aware of the benefits of technology for behavioral and therapeutic interventions, access was not always provided. This could be due to several complex factors, including the fear of expensive technology being broken easily by children or socio-cultural norms that seem to place the responsibility of upbringing on to mothers without information or economic support. Researchers at Tamana, and I, were therefore not only designing technology for specific skills, but we were also spreading awareness of and advocating for technology usage for individuals with developmental disabilities. Our observations and experiences are also formulated into several guidelines mentioned in section 6.1 and 6.2, including those focusing on addressing fears towards technology, designing for socio-technical aspirations, and promoting social inclusion thereby empowering the individuals and different stakeholders involved.

At Deepalaya, children come from underprivileged families, and the earning member, usually the father or another male member, works in nearby factories. There is a general consensus in the community on the benefits of education (computer and English language skills), especially for employment prospects (Kam, Mathur, Kumar, & Cann, 2009; Pal et al., 2007). Several students plan to continue on to higher education, which provides opportunities for a stable government job. However, even within Deepalaya, there is an observable difference between parental attitudes towards education for a girl vs. a boy. For instance, the girl is expected to get married, raise children, and take care of her family. Thus, the parents argue, there is no need to spend on her education beyond a certain age (Pal et al., 2007). Or, as seen in Deepalaya computer courses, parents are reluctant to send their teenage daughters for training because it is difficult for her to be independent or for them to offer her such freedom. These
social challenges are only overcome when technology is introduced to the community as a whole, where the ownership is communal and benefits for each individual evident (Karusala et al., 2017).

During the work with Deepalaya, it was evident that technology was well accepted by the children and teachers, given their own use of smartboards in the classrooms. The challenges were at two levels: first, the economic aspects of technology introduction and its long-term adoption (for instance, how new technology, and its regular maintenance, can be paid for). Second, understanding how children can be united with their global peers towards common learning goals once the technology is available (in this case, access to computers and the internet). The work presented in this dissertation focused on the second aspect. It also relates to increasing aspiration of Indian parents towards their children achieving expert computer and English literacy for improving their future career prospects. CityCompass provides a means to combine computer and English literacy in a playful manner and the Bollywood Method improves cross-cultural collaborations. Together, they opened a way to connect schoolchildren from varying socio-economic backgrounds across the world.

The research work itself is an example of cross-cultural multidisciplinary collaborations: with TAUCHI, IBM Research India, Tamana, and Deepalaya. Different stakeholders, researchers, teachers, therapists, special educators, and sometimes also children, have contributed towards the design and development of the application and in the analysis of the results. The inclusive nature of this work, as is common with HCI4D research, is essential when considering sustainability, that is, long-term technology adoption by the target user groups. Although the studies conducted for this research provide a snapshot on the benefits of emerging technology for underserved children, their collaborative and inclusive approach implies that the findings lay the ground work for understanding and designing for sustainable outcomes. Moreover, the guidelines provide a broad view of the socio-cultural and socio-technical challenges and opportunities towards technology introduction and acceptance, and consequently, adoption.

Future work

Based on the outcomes of the work presented in this dissertation, there are several opportunities for future work. First, given the importance of the different stakeholders towards technology acceptance, access, and adoption, and based on previous research in education in HCI4D (Karusala et al., 2017; Pal et al., 2007; Vishwanath et al., 2016), it is essential to further explore parental perspectives towards education technology. Second, there needs to be a strong focus on collaborative and inclusive educational applications and interactions.
Parents in all cultures have certain desires and aspirations towards the future of their child, some of which are achievable through technology. Moreover, the motivation for technology adoption for children is driven by the different stakeholders that surround them, including their parents and teachers. For instance, the applications designed with Tamana as a part of this dissertation are available to the centers. However, a teacher or specialist must incorporate them into their lesson plans for the children to get access. This means the specialist at the schools should be motivated to use them or the administration must insist on using them. Therefore, understanding the varying agendas, goals, and aspirations of the stakeholders and the children can provide potential clues towards sustainable technology adoption. To understand the challenges towards technology adoption, even when it is clearly beneficial for the children, I started conducting interviews with different stakeholders at Tamana and Deepalaya and at Dharavi in Mumbai (Sharma et al., 2017). Dharavi is one of Asia’s largest slums with an estimated population of one million people from different parts of India living there since the 1960s. Similar to the community near Deepalaya, Dharavi also consists of people with varying levels of literacy: in their native language, in English, and with technology. The outcomes of the work are expected to shed light on designing sustainable technology intervention for the education of underserved children.

Presently, the potential of connecting children from Tamana, Deepalaya, and schools in Finland has not been fully explored. Studies with Finnish children attending a public school in Tampere using CityCompass to collaborate with an Indian researcher are underway. Furthermore, the potential of connecting a special school in Tampere to Tamana is also being explored. However, the context will be truly inclusive when children with developmental disabilities are socially integrated with their typically developed peers. Social inclusion with peers in schools around the world and in general society is an important next step for this research, starting with India and Finland.
7 Conclusion

This dissertation presented the design, development, and evaluation of collaborative applications for underserved children in India. The work was carried out with NGOs in New Delhi: Tamana and Deepalaya. Tamana provides therapeutic, vocational, and formal education to children with developmental disabilities. Deepalaya provides remedial, vocational, and formal education to underprivileged children from informal, urban settlements. Previous research has established the benefits of interactive applications for both groups of children. However, much of this research is focused on individual interactions and within the developed work. Moreover, the underserved context provides additional challenges towards access to technology, such as lack of resources, lack of information and awareness, and other socio-cultural barriers.

The main goal of this research was to answer the research question:

How to design and introduce collaborative educational applications for social, life, and career skills to children who are currently underserved by technology due to lack of informational, socio-cultural, or economical resources?

This was divided into three sub-questions (SQ1-3) and addressed through exploratory user studies where:

SQ1. Children with autism collaborated with a typically developed individual to experience joint attention, or shared attention, towards virtual balloons using verbal or nonverbal cues, which promoted social inclusion and interactions.

SQ2. Children with developmental disabilities practiced the life skill of buying groceries form a local store in a safe and controlled environment by breaking down complex tasks into smaller
subtasks, where the learning with the virtual store was translated to a real-world scenario.

SQ3. *Children from informal, urban settlements* in Delhi collaborated on a one-on-one basis with a Finnish researcher to complete an online task, which incorporated English conversational skills, computer know-how, and cross-cultural interactions.

Results from user studies with a total of 107 children were then distilled into guidelines to help overcome the potential challenges towards introduction, access, and adoption of emerging technologies for children within the underserved context. The guidelines focused on three distinct themes:

- Reducing barriers towards technology introduction and acceptance by considering the community as a whole, empowering the different stakeholders, addressing their fears and concerns, and designing for their socio-technical aspirations.

- Designing applications to promote self-efficacy and agency, conducting user studies in way that respects participant privacy and the research environment, supports diversity and inclusion, and prioritizes participants’ interactional and instructional needs when there are conflicts with research procedures.

- Designing interactive applications for children with developmental disabilities such that they support collaboration and social inclusion, provide multimodal feedback and rewards, are structured and offer opportunities to grow, and assist in understanding real-world scenarios.

The guidelines contribute to the research with underserved children, focusing on how to design, deploy, and evaluate interactive applications within a school environment. This dissertation aims to provide fellow researchers, designers, and practitioners an understanding and a way to approach designing and deploying technology for underserved children across the world.
References


doi:10.1145/1979742.1979841

Bernardini, S., Porayska-Pomsta, K., & Smith, T. J. (2014). ECHOES: An 
intelligent serious game for fostering social communication in children with 
doi:10.1016/j.ins.2013.10.027

(2015). Designing motion-based activities to engage students with autism in 
classroom settings. In Proceedings of the 14th International Conference on 
Interaction Design and Children (IDC’14), 69–78. New York, NY, USA: ACM.
doi:10.1145/2771839.2771847

Bhuiyan, M., & Picking, R. (2009). Gesture-controlled user interfaces, what have 
we done and what’s next? In Proceedings of the Fifth Collaborative 
Research Symposium on Security, E-Learning, Internet and Networking 
(SEIN 2009), Darmstadt, Germany, 25–29.

experimental study of Hong Kong students. Social Research Center, Chinese 
University of Hong Kong.

doi:10.1145/2049536.2049542

with Kinect to Students with Cognitive Disabilities: Improving all Motor 
Skills. Procedia - Social and Behavioral Sciences, 237 (June 2016), 1148– 
1151. doi:10.1016/j.sbspro.2017.02.170

Brignull, H., & Rogers, Y. (2003). Enticing people to interact with large public 
displays in public spaces. In Proceeding of the IFIP Conference on Human- 
Computer Interaction (INTERACT’03), 17–24.

FroggyBobby: An exergame to support children with motor problems 
practicing motor coordination exercises during therapeutic interventions. 
Computers in Human Behavior, 71, 479–498. doi:10.1016/j.chb.2015.05.055


and Avatar Technology (SLTAT).


0919.1014


doi:10.1016/j.system.2015.04.001

conflict: an updated face-negotiation theory. International Journal of

doi:10.1561/1100000021

Toyama, K. (2018). From needs to aspirations in information technology for
doi:10.1080/02681102.2017.1310713

John Wiley & Sons.

Kinect-based self-modeling intervention for children with autism. In
Proceedings of the IEEE International Conference on Multimedia and Expo
(ICME ’15), 1–6. doi:10.1109/ICME.2015.7177518
Reprinted Figure 3 (right), with permission.

sparking civic discourse by a public interactive poll visualization. In
Proceedings of the 17th ACM conference on Computer supported
cooperative work & social computing (CSCW’14), 1323–1332. New York,
NY, USA: ACM. doi:10.1145/2531602.2531639

Community-Moderated Voice Forum in Rural India. In Proceedings of the
SIGCHI Conference on Human Factors in Computing Systems (CHI ’15),

of culture and implications for intercultural collaborative online learning.
Intercultural Collaboration, 260–275. doi:10.1007/978-3-540-74000-1_20

Vellanki, P., Greenhill, S., Duong, T., Phung, D., Venkatesh, S., Godwin, J.,
for India. In Proceedings of the 28th Australian Conference on Computer-
Human Interaction (OzCHI’16), 618–622. New York, NY, USA: ACM.
doi:10.1145/3010915.3011007


103

Promoting Joint Attention with Computer Supported Collaboration in Children with Autism

Sumita Sharma¹, Saurabh Srivastava²*, Krishnaveni Achary³, Blessin Varkey³, Tomi Heimonen¹, Jaakko Hakulinen¹, Markku Turunen¹, and Nitendra Rajput⁴

¹University of Tampere, Tampere, Finland
²Xerox Research Centre India, Bangalore, India
³Autism Research Center, Tamana School of Hope, New Delhi, India
⁴IBM Research, New Delhi, India

{firstname.lastname, sumita.s.sharma}@sis.uta.fi, saurabh.srivastava@xerox.com,
{krishnaveni.achary, blessinvarkey}@gmail.com, rnitendra@in.ibm.com

ABSTRACT
There exists mounting evidence in favor of computer supported autism interventions at the individual level. However, the potential benefits of using computer supported collaboration to encourage social interactions between individuals with autism and typically developed individuals are underexplored, particularly in developing regions. We present an exploratory study of a collaborative gesture-based application, Balloons. The application encourages joint attention, which is defined as the shared attention between two individuals towards the same object. Using mixed methods, we evaluated Balloons for three weeks in New Delhi with ten medium-low functioning autistic children. Our findings suggest that employing CSC interventions for children with autism in India provide (a) observable improvements in social interaction with typically developed peers, (b) the opportunity to customize and individualize intervention to cater to a large spectrum of children and (c) the potential opportunity of reducing fears of certain objects.

Author Keywords
Computer Supported Collaborative Work; Gesture-based Interaction; Joint Attention; Autism Spectrum Disorder

ACM Classification Keywords
H.5.3 [Information Interfaces]: Group and Organization Interfaces - Computer-supported cooperative work.

General Terms
Design and Human Factors

INTRODUCTION
Autism is highly individualized and hence, the term Autism Spectrum Disorder (ASD) is used to represent the varying levels of behavioral, cognitive and social challenges the individual faces. Previous research has established that computer supported interventions provide a controlled and safe environment for individuals on the autism spectrum [4 and 21]. However, research on computer supported autism interventions is primarily focused at either the individual level [5, 14, 17, 21, 26], asynchronous online social communication [7] or human-robot social interaction [6, 27, 28]. Furthermore, much of the research is concentrated in the developed world, neglecting emerging economies, where Autism is largely misunderstood. Studies show that in India, autism diagnosis is affected by inherent social taboos regarding mental illnesses, the lack of awareness, the lack of relevant information for parents and first level local pediatricians, and the limited number of schools and therapeutic centers [10 and 18]. For children on the autism spectrum early interventions are critical for maximizing outcomes [11] and thus, as soon as a child is diagnosed, they require immediate intensive and individualized interventions.

In this paper, we draw attention to the use of computer-supported collaboration (CSC) to encourage social interaction between medium-low functioning autistic children and typically developed individuals. Medium-low functioning autistic children have a limited vocabulary, low motivation for social interaction and are prone to displaying socially isolating repetitive behavior. Our goal was to study ways to facilitate the adoption of the critical skill of joint attention [8], defined as the shared focus of two individuals on the same object. Working with a special school in New Delhi, India, we explored the potential benefits of using gesture-based CSC within the classroom environment. Although the school is already working with tablet-based interventions for autism, the TOBY Playpad [31], they have found that tablets themselves tend to limit the collaborative nature of social interactions and confine the children to uncomfortable stationary positions. Additionally, the school

* This work was done while Saurabh was affiliated with IBM Research.
lacked interventions focused on joint attention. To address these issues, we designed, developed, and evaluated a gesture-based application to support joint attention, called Balloons, which encourages two individuals to simultaneously select an object. Our work started with a user-centered design process of the application wherein researchers held group discussions with school specialists consisting of educators, occupational therapists, speech therapists and psychologists. An exploratory user study, using mixed methods analysis, was carried out over three weeks with ten medium-low functioning autistic children collaborating with typically developed individuals.

Our study is one of a kind at two levels. First, we employed gesture-based interaction for underserved Indian children on the autism spectrum. Second, we applied gesture-based CSC for joint attention between medium-low functioning children with autism and typically developed individuals. To our knowledge, the only study close to our work is reported by Bauminger-Zviely et al. [5] who looked at socially collaborative interventions, but using multi-mice or multi-touch interactions and for collaboration among high (not medium-low) functioning autistic children in Israel.

In our study we observed that the social and cognitive challenges faced by medium-low functioning Indian children with autism are similar to those documented by Bartoli et al. [3 and 4] from a developed world context. Our findings suggest that by supporting collaborative interaction paradigms, children on the autism spectrum can be introduced to subtle social interactions. This builds confidence and independence, and provides a mechanism to be integrated with typically developed peers. We propose tapping into the potential of using virtual objects and environments to reduce fears or phobias of individuals with autism. Furthermore, specialists can use these applications in innovative ways for promoting social collaborations, and motor and verbal skills. For example, by using target pointing to improve postural stability and including different objects to teach vocabulary.

Specific to the Indian context, it was interesting to observe the motivation to customize and individualize a simple gesture-based application to teach several different skills to several different children on the autism spectrum. This resonates with the observable Indian attitude of jugaad or innovative problem solving with limited resources [32]. The school educators also played the role of social mediators helping promote technology acceptance in the community.

In this paper, we first situate our research with the related work in this domain, followed by a description of our user-centered design process, and the Balloons application and its interaction. Next, we present our user study at the school including the evaluation methodology. Then we present our results and extend our observations to bring out several common themes as identified by the interviews with the school specialists. We conclude by discussing the implications of our results and findings to gesture-based CSC research in autism.

RELATED WORK

We describe the previous research by first understanding the role of joint attention in autism and associated CSC interventions. Second, we present the technological interventions focused on individual interaction and not collaboration. We then highlight the opportunities for employing gesture-based interaction for joint attention interventions in CSC applications. What must be noted here is that current research on technological interventions for children with autism is spread across the developed world, with little to no reports from the developing world.

Joint Attention and Autism

Joint attention pays a pivotal role in early autism detection and is linked to language acquisition and social interaction in the later stages of the neurological development of a child [8]. Joint attention is defined as the social interaction over a common goal or object relating to shared attention and experience, regarded as a critical skill for children on the autism spectrum. In fact, lack of joint attention plays a crucial role in diagnosing autism in infants [13]. Befittingly, interventions for young children focus on improving joint attention and shared experiences between children with autism and their caregivers [23]. These interventions are largely video recorded and later analyzed by the therapist to extrapolate the object of attention by observing the child’s gaze and attention [16, 22], which is laborious. The focus in technology-driven interventions has shifted towards employing computer-supported instructions [14] to individuals with autism, largely ignoring CSC for joint attention between individuals with autism and typically developed individuals.

In the last 15 years significant research has focused on joint attention mediated by virtual conversational agents [1 and 2] or robots [5, 27, and 28]. In the Aurora project [27, 28], researchers studied social interaction, mediated by robot-toys, between children with autism and their care-givers or therapists. They also studied how the appearance of robots affected the child’s interaction, specifically the less human features (clothes, face) the robot had, the more appealing it was. Although Bekele et al. [6] developed an adaptive and individualized robot to encourage social interaction, using speech and gaze, between children with autism and the robot; it was only evaluated with typically developed individuals. ECHOES is a serious game with intelligent virtual characters for encouraging social communication employing gaze and large touch screens [1, 2]. However, it also remains to be studied if interaction with virtual characters promotes social communication between children with autism and typically developed individuals.

Bauminger-Zviely et al. [5], developed games for collaboration among high functioning autistic children using multi-mice and multi-touch interfaces. Their results showed that children with autism gradually improved their
social conversation, joint collaborations, and problem-solving skills strengthening the case for computer supported collaboration for joint attention interventions. However, they do not consider bodily interaction and the inherent benefits it offers for children with autism.

Technology for Individual Interactions
Technology can be used to provide a controlled, safe and stimulating environment for children with autism who are known to face challenges in social interaction, communication, and imagination. Previous research also strongly suggests that information and entertainment technologies, including tablets, mobile phones, TV and video games, are appealing to autistic children [14, 21]. Consequently, emerging HCI research in autism is heading towards autism interventions employing various interaction technologies. For example, The TOBY (Therapy Outcomes by You) Playpad is an iPad educational intervention for autistic children, which “aims to empower parents to commence early intervention to maximize their child’s development” [31]. It provides touch-based interactive games focused on improving cognitive abilities such as attention, memory, object recognition, and learning of complex activities by imitation using videos.

As stated by Keay-Bright [17], computer-based learning environments as a medium help reduce the multi-sensory distractions occurring in real world, which can induce anxiety in many people with autism and may hinder their participation in social communication. Her ReacTickles software, popularly used with the Smart™ Interactive Whiteboard, promotes relaxation in children with autism by spontaneous playful activities. Moreover, Ringland et al. [26] focused on augmenting existing physical sensory integration therapies by employing tangible interaction using rubber balls. Their application, SensoryPaint, was found to improve body-awareness and proprioception in children with autism. Interestingly, both SensoryPaint and ReacTickles support the use of bodily interaction to provide more kinesthetic forms of expressions for children with autism. However, these applications focused on individual interaction and do not fully utilize the potential of computer supported social interaction between children with autism and typically developed individuals.

Gesture-based Applications for Autism
With affordable motion tracking technologies, such as the Microsoft Kinect, gaining momentum in the educational space, ongoing autism research is also moving towards immersive and engaging gesture interactions [3, 4, 9, 11, 25]. There is a strong body of research by Bartoli et al. [3, 4], the Lakeside Center for Autism², MEDIATE [24, 25] and Kinems³ providing evidence of the potential benefits of gesture interaction for therapeutic and learning experiences. We discuss their work to highlight the extensive benefits of gesture-based interaction for children with autism, which are extendable to CSC interventions.

The Lakeside Center for Autism in Washington State, US, started a Kinetix Academy in 2012 to develop applications for children with autism focusing on motor, speech, cognitive, and social skills for educational and therapeutic interventions. MEDIATE [24, 25] is a multisensory interactive environment that “generates real time visual, aural and vibrotactile stimuli” for low functioning autistic children. Children are allowed to explore the spaces at their own pace and there is no time limit to the interaction. Results from user studies, conducted in London, Hilversum (Netherlands), Barcelona, and Portsmouth have shown a high acceptance of its interactive space by autistic children, who are able to creatively express themselves using the several multimodal applications presented in the space.

A more recent online community, Kinems, spread across the US, Greece, and Netherlands develops Kineckt-based games to improve cognitive abilities of children with learning disabilities including hand-eye coordination, memory, attention, and basic problem solving. Their findings suggest that even though kinesthetic interaction induces fatigue, it provides an immersive and engaging learning environment that can help children concentrate. There is also research on gesture-based games to match visual facial expressions [9] using existing Kinect sports games to improve hand-eye coordination, attention and focus [1], and gesture recognition to detect repetitive behavior or tantrums [12].

Furthermore, research by Bartoli et al. [4] provides design guidelines for gesture-based playful interactions based on their ongoing research with children in autism in Italy. They divide their guidelines as general or goal specific focusing on motor, cognitive, and social skills which are important considerations for children classified with medium-low functioning autism.

COLLABORATIVE BALLOON APPLICATION
Our collaborative balloon application focuses on promoting joint attention between medium-low functioning autistic children and typically developed individuals. Its users have the task of collaboratively selecting one of three balloons by pointing to it. This necessitates the need for social interaction between the autistic participant and their typically developed team mate. The use of a CSC system reduces the social awkwardness of face to face social communication or making direct eye contact, a defining characteristic of autism.

Design Process
Our application was designed through a user-centered design approach involving twenty-three participants: nineteen school specialists, three parents and two high functioning autistic children. In order to understand the process of conducting educational interventions, two researchers spent the first week at the school becoming

AUTHOR’S VERSION

---

² http://lakesideautism.com/
³ http://kinems.com/
familiar with the ongoing interventions and therapies. They observed various specialists conduct group sessions and visited each of the therapy labs in the school including the speech therapy rooms, occupational therapy rooms, psychologist’s rooms, a computer lab, the TOBY Playpad lab, and the dance halls.

After this first week, semi-structured interviews were conducted over the course of two weeks with nineteen specialists: twelve educators, two occupational therapists, two psychologists, two speech therapists and also three parents who attended the school with their child. Two of the twelve educators were responsible for technology-based intervention for the school and were involved in the TOBY Playpad [31] interventions. These two educators worked very closely with the researchers; they were present in every discussion, moderated the evaluation sessions, and are also coauthors of this work. The focus of these interviews was to uncover specialists’ challenges with their own specific therapy sessions and their suggestions on how to overcome them. The interviews usually started with a specialist demonstrating their usual session protocol, followed by in-depth discussions about the specific therapy, its outcomes and challenges, and how to introduce technological aids. The specialist also shared their own intervention design guidelines as compiled over several years of hands-on work with the children.

Based on the common themes that emerged from the interviews, we conducted two focus group discussions during the fourth week. We started by consolidating a list of application requirements, and then built on them to design and develop our application’s interaction, interface, and intensio. These discussions included the school psychologists and the two educators who work with the TOBY Playpad. It was during these discussions that the familiarity of balloons as a rewarding object for the children, taken from the TOBY Playpad observations, was mentioned several times. This led us to incorporate balloons in our own application. After developing the application, we further interacted with two high functioning autistic children from the school to pilot test our application.

We summarize the school’s prime aim as being able to develop self-sustainable life skills in children with autism to allow them to function well independently. Children with autism face several challenges including personal hygiene, concepts of collaborative gameplay and team work, and a lack of motivation for social interaction. The children occasionally display repetitive behavior when stressed or uncomfortable, and even while enjoying a certain activity. This type of behavior is socially isolating for the child and requires special attention. The specialists mentioned the success of occupational therapies involving bodily movement for self-expression and exercises to pacify the children behaving hyperactively. We next present our consolidated set of application requirements.

**Application Requirements**

A. Designing for improvement of social skills of medium-low functioning autistic children in the school who have limited verbal communication. The skill was identified as joint attention because of its importance in social interaction for autistic children and lack of such a joint attention focused technology intervention currently at the school.

B. Providing adequate rewards to encourage children to interact, regardless of their performance, for example, applause from the system.

C. Providing a well-defined start and end of a session in order to allow the children to understand when a task is over.

D. Encouraging socially acceptable gestures and a simple pointing gesture was decided upon for the application. Using gestures such as kicking or jumping were thought to possibly provide encouragement to kick and jump outside of the session, which can be dangerous and socially isolating.

E. Providing multimodal feedback, both visual and auditory, as the children have varying levels of visual and auditory preferences, sensitivities, and capabilities.

F. Include colorful graphics with smooth visual movements and illustrations to captivate attention for longer time spans.

G. The interaction should be empowering by involving the child in decision making.

**Gesture Interaction Design**

Our CSC application focuses on encouraging medium-low functioning autistic children to explore proto-declarative pointing [15] leading to shared experiences by virtue of joint attention (requirement A). The application interface consists of three different colored balloons (requirement F). A player’s hands are represented by onscreen hand cursors that move accordingly (requirement G). A snapshot of the game with one player’s right hand is shown in Figure 1.

![Figure 1: Application screen with a participant’s right hand pointing near the blue balloon](image)

The application requires a two member team to collaboratively select the same virtual balloon out of the three possible options, further encouraging the need for social interaction, verbal or non-verbal, between the team members. Balloons are selected by pointing for a fixed dwell time (requirement D). If one player member selects a
balloon, a star is shown on the screen as a reward. If both team members collaboratively select the same balloon, a rainbow that grows in size is shown on the screen with pleasant background music (requirements B and E). The rainbow and star are shown in Figure 2.

The balloon application requires two players to simultaneously select one balloon (requirement G) and after trying the application with the school specialist and two high functioning autistic children (who were not a part of our subsequent user study) the dwell time for selection was set to three seconds. To increase the level of difficulty, the balloons would slowly fly upwards after five minutes of continuous interaction.

Figure 2: A star and a rainbow as visual rewards

![Figure 2: A star and a rainbow as visual rewards](image)

The evaluation core logic of the Kinect service is a thin client over the Microsoft Kinect SDK that connects to the core logic over TCP socket. Microsoft Kinect is used to track a user’s hand with respect to his body within a 3D pointing area, called the physical interaction zone [20]. This physical interaction zone is relative to the user’s hand, and not the display, making it more flexible to point at larger distances (Figure 4).

All audio and graphical content is rendered using the Panda 3D\(^4\) graphics engine. The core logic is a Python based application that consists of an input/output management module, handling all the inter-process communication and acting as an interface between the core logic, the Kinect, and the graphics engine.

```python
import numpy as np
import panda3d

class Ballons:
    def __init__(self):
        self.balloons = np.array([...])
        self.graphics_engine = panda3d.engine

    def evaluate(self):
        pass
```

Figure 4: Physical interaction zone area, in blue, as seen from the (a) front (b) side (c) bird’s-eye view (adapted from [20])

![Figure 4: Physical interaction zone area, in blue, as seen from the (a) front (b) side (c) bird’s-eye view](image)

The application has a predefined active area where the system starts responding to a user’s gestures (Figure 3). This active area is 4 m x 1.5 m, 1 m away from the center of the Kinect. A game starts when the Kinect system detects at least one player in the active area and plays a welcome message followed by a short music clip and a black background screen fades out to a white application screen as shown in Figure 1 (requirement C). While a player is in the active area, the system responds to his gestures. The application allows for a maximum of two players at any given time. When there is no longer a player in the active area, the application screen goes back to its initial black screen and an applause audio clip is played (requirement B).

EVALUATION METHODOLOGY

We evaluated Balloons with ten medium-low functioning autistic children from the school over a course of three weeks, with two to four sessions per week. The application was installed in one of the school’s classrooms, which is commonly used by the participants in order to provide a familiar environment. The evaluations were moderated by the two school specialists who are a part of our research team. There were no external researchers present during these sessions to eliminate any unnecessary stress and discomfort for the children caused by an unfamiliar presence. The absence of external observers was compensated by the specialists being better equipped to understand the participants’ emotional response. For

\(^4\) https://www.panda3d.org/

AUTHOR’S VERSION
example, they understood whether a participant’s tantrum is because of excitement, boredom, or a need to visit the restroom. A session was held before the evaluations where they were briefed on the observational requirements and provided easy to fill in data sheets with a behavioral signals table, adapted from Bartoli et al. [3]. The behavioral observations comprised of participant specific behavior to indicate positive emotions, frustrations or distress, need for assistance and issues in interaction with Balloons. Furthermore, the specialists were extensively interviewed after the evaluations.

As mentioned in the design process, we decided to use balloons because the children were already familiar with them from the TOBY Playpad sessions. This removed the complexity of introducing a new object and eliminated the effects of joint attention motivated by interaction with a novel or unfamiliar object. During the evaluations, our ten participants were exempt from all similar activities and applications that targeted social skills, thus these sessions were conducted as a separate activity.

To understand if the children are able to transfer learning of joint attention interactions from one context to another, we devised the physical balloon test. Children with autism find it difficult to draw parallels between two similar contexts, for example, what is suitable behavior in the occupational lab such as jumping, is not suitable in other classrooms. The physical balloon tests had tasks similar to the virtual application but with real balloons and in a different classroom, in order to understand whether joint attention skills carried over to the real world from the virtual.

**Ethical Considerations**

Parents or legal guardians provided consent for participation for all the children involved in our study. The school insisted that all children be provided the potential benefits of the application and no child should be left behind. To honor their request, we did not establish a control group for our user studies. The absence of a control group steered us away from focusing on the potential learning gains from our application. Instead we chose to understand the challenges and opportunities of introducing gesture-based CSC interventions in a classroom, and as explained later, this provided us with a far more interesting perspective. We made the application available to the school after the evaluation.

**Participants**

Studies were conducted with 10 autistic children (6 male, 4 female) diagnosed by the Autism Diagnostic Observation Schedule (ADOS) [19]. These ten children also met the criteria for autism on the Childhood Autism Rating Scale (CARS) [30] (M=34, SD=3) and the Social communication Questionnaire (SCQ) [29] (M=48, SD=13). Seven of the ten children have a vocabulary of 100-200 words while the other three were considered non-verbal communicators with a vocabulary of less than 50 words. All children were categorized as medium-low functioning autistic children by the school therapist as they had low motivation for social interaction and communication (both verbal and non-verbal), and displayed various isolating behaviors and interests. The participants were able to understand simple instructions, and displayed repetitive behaviors, attention, mood behaviors, and lacked the social skill of joint attention. None of our participants had any prior experience with gesture-based interactions, gaming or otherwise.

**Procedure**

The procedure consisted of three phases; phases I and III were carried out with actual balloons for a pre and post-trial analysis and phase II with the Balloons application. This also helped us understand whether the learning from a virtual application is translated to the physical world. The ‘autistic participants’, referred to as ‘participants’ from now onwards, were asked to choose a typically developed peer or a school specialist as a team member, depending on their comfort level. Although the causal effect between participants’ familiarity with their teammate and their improved task performance is difficult to measure, we must note here that our participants lacked joint attention skills regardless of their familiarity with their team mate, which suggests that any familiarity effect was minimal. Conducting the evaluations in a known environment and with a known teammate reduced the participants’ social discomfort allowing for serendipitous interactions.

![Figure 5: A participant (center) and his teammate (left)](image)

**Phases I and III: Physical Balloon Test**

In a physical balloon test, actual balloons of different colors were arranged on the school’s wall as shown in Figure 5.

The task included two steps:

1. A participant’s typically developed team mate had to select one of the colored balloons on the wall and point towards it, standing at a distance.

2. The participant had to then go and touch that balloon.

![Figure 6: A participant playing Balloons](image)
For collaboration, the team mate used both physical prompting (pointing) and verbal prompting (saying the color). This task was repeated three times with different colored balloons. The pre-trial physical balloon tests assisted in understanding a participant’s ability for social interaction and joint attention by pointing. The phase I participants then also took part in the phases II and III.

**Phase II: Balloons Application**

Phase II consisted of the virtual balloon application and the Kinect setup was installed in one of the classrooms of the school as shown in Figure 6. Phase II evaluation consisted of a maximum of eleven sessions (n=4) and a minimum of three sessions (n=1) per participant during the three week period (M=8, SD=3). This large variation was due to two participants who did not attend the school regularly during the evaluations. The ten participants spent an average of 1 minute and 46 seconds per session (SD=24 seconds).

**Data Collection**

During the phase II trials, we collected system logs for the number of balloons selected, specialist observations, and behavioral responses of the participants. These behavioral responses consisted of commonly expressed gestures or habits by the participants to indicate positive emotions (such as smiling or clapping), frustrations or distress (loss of attention, asocial or inappropriate gesturing), need for assistance (verbal or physical pacifying) and issues with system interactions [3]. When working with participants with autism, these behavioral signals provide a way to understand their experience as whole and are also a means of communication with their care-giver and therapists.

In phases I and III, the task completion time and outcomes were collected. The school specialists, who conducted the sessions, were later interviewed to understand in detail the various themes that emerged from their observations. The interviews were semi-structured and focused on understanding the participants’ and specialists’ experience with Balloons.

**RESULTS**

During the three evaluation phases, the moderators observed changes in participants’ interaction experiences with the task and our application. During phase I, the participants familiarized themselves with the task, and they possibly carried over the understanding of the task from phase I to phase II. During phase II, the participants first learnt how to use the application and then collaborate with their team mates, which carried forward to phase III where we noticed an improvement in average balloon selection time. In phase I (pre-trial) the average time to select the first balloon collaboratively was 66 seconds (SD = 65). In phase III (post-trial), average time to select the first balloon collaboratively was 7 seconds (SD = 5). This shows a promising translation of learning from the Balloons application to the real world scenario.

Interestingly, the school specialists found two ways to use Balloons for joint attention. The first innovative method materialized upon need and consisted of the specialist standing behind the participant to guide the interaction, as show in Figure 7. This way, the specialist was able to collaboratively point and guide at the same time, although the system would recognize them as a single user and show the star as a reward. The second was as intended, by standing next to the participant and collaboratively selecting the same balloon. In this case the system would recognize two users and a rainbow would be shown as the reward. The sessions started with the first method of collaboration and gradually moved towards the second. We also observed that this transition from the first method to the second started with the blue balloon, probably because of its central position. Out of a total of 280 successful balloons collaboratively selected in phase II, 146 (52%) were by the first method (star as the reward) and the remaining 134 (47%) by the second. This shows a clear potential for the application in improving joint attention and shared experiences, from instructional-based collaboration to collaborating as individuals.

![Figure 7: A specialist guides the child's interaction](image)

![Figure 8: Average number of balloons selected per minute for each session in Phase II](image)

Figure 8 shows the average number of balloons selected per minute in each session for the ten participants. It must be noted here that the slight dip in the 5th and 6th sessions indicates the transition between the two modes of collaboration. Although we can see an overall improvement in the average number of joint attention collaborations between the first and the last session; actual improvements varied across participants. This can be seen by the increasingly varied SD from sessions 8 to 11. We attribute this to (a) the number of participants who did not complete
all 11 sessions (n=6) and (b) the variability of improvements commonly expected within the autistic spectrum.

Looking at the individual performance of participants who attended 8 or more sessions in phase II (n=5) in Figure 9, it is evident that participants P1 and P5 do not show any improvement while participants P2, P3 and P4 do. We purport that given enough sessions, all our participants would have benefited from the phase II intervention. We discuss next the implications of our results and the challenges and opportunities of our study.

![Figure 9: Individual Performance in Phase II of Participants who attended 8 sessions or more (P1-P5)](image)

**DISCUSSION**

Our results suggest that using gesture-based interaction for CSC between medium-low functioning autistic children and typically developed individuals is a promising way forward. Although the lack of a control group steers us away from strongly claiming our potential learning benefits, our behavioral observations and post-evaluation interviews reveal several challenges and the opportunities of our intervention. Collaborative gesture-based applications can utilize the appeal of visual stimuli and provide engaging classroom activities that allow caregivers to control the learning environment, and the possibility to improve postural stability; findings that resonate with previous research from the developed world [3, 24]. Additionally, these applications provide a mechanism to help children overcome fears of specific object. These user specific findings seem to be culturally independent for medium-low autistic children using gesture interaction for joint attention. Thus, we believe that developed world interventions can largely benefit children with autism in India, and vice versa.

The benefits of using CSC for promoting joint attention, as uncovered from our study, include motivating the educators to employ innovative method for social collaboration and overcoming the limitations of the tablet-based interventions. While these benefits seem to be extendable to the developed world, further research is required to support it. Furthermore, we experienced a high expectation of technology from the school educators and caregivers, due to the Indian need of using a single resource to try to solve everything for everyone. We also note the collaborative nature of our entire study; from designing the application with the educators to empowering them as social and technology agents of change. Next, we expand on these findings and discuss the benefits of using CSC applications for children with Autism.

**Behavioral Observations and Post Evaluation Interviews**

**Overcoming Fear of Balloons**

One of the participants was scared of balloons, and as observed during phase I he would not touch the physical balloon but stand next to it with his ears covered. However, during phase II evaluations, he realized that the virtual balloons did not pop loudly and instead have pleasant rewards. He began to interact with the application. This was extremely exciting for the specialists, who were able to use our application to reduce his fear of balloons, and they expressed their desire to apply this to other children who might fear other specific objects. Using a virtual application to overcome phobias in children with autism is a relatively unexplored, with the exception of the Blueroom project.

**Rewards and Feedback**

Bartoli et al. [4] emphasize motivating children with autism by using appropriate rewards and feedback. We also noticed that audio feedback (applause) was successful in making the participant feel rewarded regardless of their performance as several participants would also start clapping. The specialists suggested the use of culturally specific objects or the participant’s current object of affection for more effective rewards. From their experience with the tablet based interventions, they found that firecrackers, emblematic of the Indian festive of Diwali, were highly fascinating for the children. However, our participants in the study, as observed with children with autism, were generally more inclined towards visual stimuli. Whether fire-crackers are a cultural reinforcement or have a more universal appeal, remains to be studied.

**Attention and Engagement**

Although the application supported movement of balloons, the configurable delay of five minutes was too long and not changeable on the fly. The specialists also noted that the three second dwell time was too quick for some participants to collaboratively select the balloon, resulting in a star and not a rainbow as the reward. This caused a loss of attention and interest. One of the participants with known behavioral emotion of walking out of the room when bored, would walk out of the classroom. Such sudden inattention and disengagement are common behavioral responses from children with autism [4].

However, one of the participants found Balloons background color change, from black (system inactive) to white (system active), very intriguing. He would purposefully move between the active and inactive area to experience this dramatic visual change, but was then unable to focus on the application tasks.

5 http://www.ncl.ac.uk/ion/news/news/blueroom/
The two school educators, who moderated the evaluations, foresee these digital interventions as a catalyst in enhancing learning outcomes while keeping the participants interested and engaged. One of them mentioned that engaging digital medium of learning encourages classroom activities, “when a student is engaged with a system, I get enough time to focus and monitor them one by one”. The other said that “in special needs school like ours, the management greatly values how we handle our students, and maintain the decorum of the class. This system is very helpful as while one student performs, other students are engaged in watching his activities.”

We observed several different behavioral attitudes of the participants; ranging from the positive emotions of excitement and happiness (clapping and hugging the specialists) to the more negative feeling of frustration (physically abusive). For instance, one of the participants repeatedly requested to play the game over and over again. This is quite a common phenomenon in autism interventions; participants become too attached to a specific object or game.

Improving Postural Stability
The specialist noticed an improvement in the hand movement of two participants as the gesture interaction required a steady hand increasing postural stability [4]. These participants were able to use the on-screen cursor to guide their hands to the balloon and with the dwell time of three seconds, encouraged them to keep their hands steady for that duration. This can potentially improve hand-eye coordination, body awareness and pointing at a target, which was highly motivating for the specialists as the benefits were quickly visible and experienced and also carried forward to real world situation (pointing with a steady hand). The specialist suggested being able to vary the balloon (target) size and dwell time (hold steady) for continuing the improvement in postural stability.

Benefits of Computer Supported Collaboration for Children with Autism
The benefits of gesture-based interaction for children with autism are well established in the developed world [3, 4, 9, 24, 25]. We discuss our findings from our collaborative gesture-based CSC application for children with autism.

Innovative Method for Collaboration
Our most positive finding is from the school specialist’s usage of balloons in two innovative ways for joint attention. They initially used the first method, i.e., standing behind the participants and helping them to select a balloon. Although this method would produce only a star as the reward, it proved to be beneficial for our participants who were all new to gesture interaction. The second method was the intended way, two players selecting the same balloon to be rewarded with the rainbow. In this way, the school specialists were able work collaboratively in teaching the child how to use the interaction by focusing on the same balloon. We find this in line with the Indian jugaad, or innovative life-hacks, a practice observed across various domains [32].

Moreover, after the evaluation period, several parents of our participants (n=3) commented on how their child mentioned a balloon game at home. This is very surprising because our participants did not have a tendency to share their experiences from the school at home. We believe this to be an extension of joint attention via sharing of experiences outside of the evaluation context, a promising outcome of our study. These parents also wanted the school to continue the rigorous Balloons trials, which is a big breakthrough in technology acceptance for the school.

Touch vs Touchless Interaction
The specialists discussed their different experiences in conducting the intervention between touch and touchless interactions. With the tablet based interventions, participants are able to serendipitously discover the close button on an application window or the device. Upon discovery, they tend to close the application or device as soon as they are bored or disengaged. With the gesture-based interaction, the specialists commented that distance between the actual hardware and the participants eliminated such disruptive behaviors and encouraged the participants to work longer. Another important observation is that participants sometimes tend to throw or swing a tablet, increasing the chances of getting hurt or hurting someone else. With a gesture-based system such behavior was not supported, much to the relief of the specialists. Furthermore, one of the specialists mentioned, “we have limited tablets with us and children have to wait to work on them, but a system like Balloons allows more students to be engaged while viewing the activity on the screen together.”

Furthermore, when the participants learn to use the touch modality for the tablet, they tend to extend this interaction to other objects and also to different applications within the tablet. A gesture-based touchless system allows for different modes of interaction, even using different parts of the body such as hands or feet, making it easier to differentiate between applications. It will be interesting to study how participants carry forward their interaction with the Balloon application to other gesture-based applications requiring different hand gestures. Although not used in the Balloons application, a gesture-based intervention could also study participants’ body posture or gait and share the data with the occupational therapists.

Technology Expectations
The school specialists expressed their desire to customize the application based on the participant’s individual skill level, for example, varying the virtual balloon’s color (to teach different colors) or the size (to teach postural stability). They also suggested including culturally specific or participant specific (participant’s current object of affection) rewards to increase motivation [4] and supporting multiple languages. For a multilingual country like India, participant’s parents prefer using a specific language at
home and wish that their child also learn the same in school.

This desire to customize an application for individual autistic participants resonates with previous research [4] and can be handled in various ways, for instance by creating user profiles for the participants. However, what is surprising is the high level of expectation, of customization on the fly, from a simple application focused on joint attention. As discussed before, when working with users from emerging markets, it has been observed that their expectations of technology can be higher because devices are usually seen as a shared resource that should cater to each user’s need [32], and there is an inherent scarcity of such resources. Technology needs to satisfy both the individualized learning goals and the stakeholders’ resource utilization goals. Thus, there is a strong desire to use one application for all children on the autism spectrum; that is, the expectation that a single technical intervention can solve everything for everybody.

**CSC Applications for Children with Autism in India**

The specialists shared their experiences with introducing technology-based educational interventions at the school. They have noticed an overall reluctance to adopt new technology-driven interventions, as was evident during the introduction of the tablet-based therapies in their school. However, this attitude has gradually changed and as we mentioned before, several parents of our participants insisted on the continual use of Balloons for their child. Many also enquired about the Kinect hardware cost and if they could use this system at home with a TV or even a mobile phone. One of the parents stated, “I am a strong believer of doing things practically. I am happy that the school has taken an initiative where children are being exposed to physical real life interventions.” This makes us believe that the urban Indian attitude is experiencing a positive shift towards technology creating a hotbed of opportunities for further research.

Our findings of participants’ behavioral responses to new technological interventions resonate with previous research [3, 24] suggesting that Indian children with medium-low autism face similar behavioral challenges to their developed world peers. We thus argue that children with autism in India can benefit from interventions available to the developed world. However, we also observed extremely high expectations of technology from the educators and caregivers of children with autism, which we purport to be culturally specific to developing countries.

**Extending Collaboration Beyond the Application**

We note here the collaborative nature of our entire study; educators from the school worked with us from its design to its publication. The study also invoked innovative responses to crucial needs; from the two methods of joint attention collaborations to the call for further individualization and customization to cater to a larger group of children with incremental learning goals. The application provided a mechanism to empower educators to conduct and plan more efficient and effective interventions, and benefit from the increasingly positive researched phenomenon of bodily interaction for children with autism. Additionally, the educators also played an important role as social mediators to promote technology acceptance among caregivers and other stakeholders. Based on the positive experiences from this study, the school has already dedicated resources to further explore gesture-based interventions for children with autism. We believe our small study has sparked a revolution in the special school in New Delhi and we hope to carry it forward to other schools around India.

**CONCLUSION**

Employing gesture interaction for children with autism has gained favor in recent years with many researchers working towards developing innovative applications to motivate and encourage autistic children to learn and interact socially. However, much of this research is focused on children from developed countries and is at the individual level, as most of the applications are not collaborative in nature. Our research examined the challenges and opportunities in introducing CSC for social interaction, focusing on joint attention, through pointing gestures for children with autism.

Our study is a first of its kind by employing gesture interaction for (a) children with Autism in India and (b) designing collaboration between children with autism and typically developed individuals. We identified several challenges in introducing computer supported social interventions in developing countries; such as catering to the high expectations of technology by providing on the fly customization and allowing for one application to be used by a spectrum of children and overcoming low technology acceptance.

Through our promising findings we encourage other researchers to actively extend their participation to the underserved community of children with autism from emerging economies. We also encourage other researchers working in this domain to design collaborative, not only individual, social interaction paradigms for children with autism. This assists them in building confidence, gaining independence and provides a mechanism to be integrated with typically developed individuals, which is the motivation behind most autism interventions.

**ACKNOWLEDGMENTS**

We express our sincere gratitude to all the children who took part in our study and also thank all the school specialists for their valuable time and guidance.

**REFERENCES**

1. Alyssa Alcorn, Helen Pain, Gnanathusharan Rajendran, Tim Smith, Oliver Lemon, Kaska Porayska-Pomsta, Mary Ellen Foster, Katerina Avramidis, Christopher Frauenberger, and Sara Bernardini. 2011. Social communication between virtual characters and


Gesture-based Interaction for Individuals with Developmental Disabilities in India

Sumita Sharma¹, Saurabh Srivastava², Krishnaveni Achary³, Blessin Varkey³, Tomi Heimonen⁴, Jaakko Hakulinen¹, Markku Turunen¹, and Nitendra Rajput⁵

¹University of Tampere, Tampere, Finland
²Xerox Research Centre India, Bangalore, India
³Autism Research Center, Tamana School of Hope, New Delhi, India
⁴University of Wisconsin–Stevens Point, Stevens Point, WI, USA
⁵IBM Research, New Delhi, India

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

ABSTRACT
Gesture-based interaction provides a multitude of benefits to individuals with disabilities, for example, enhancing social, motor and cognitive skills. However, applications that encourage self-efficacy by promoting a life-skill through simulations of real world scenarios are largely missing. We explore the benefits of using a gesture-based application for individuals with developmental disabilities. The context is a special school in New Delhi, Nai Disha, where we designed and developed an application, Kirana, that integrates arithmetic and social interaction to teach purchasing of items from a local grocery store. In our study, 18 participants with developmental disabilities, previously unable to visit a grocery store, used Kirana for three weeks. Our results indicate that gesture-based applications can teach a life skill and enable self-efficacy for individuals with developmental disabilities by breaking down complex tasks that require social, mathematical and decision-making skills.

Keywords
Gesture-based interaction; Developing countries; Individuals with developmental disabilities

1. INTRODUCTION
Individuals with developmental disabilities require personalized care when using conventional educational methods. They face challenges in learning skills that promote self-efficacy – including the danger of being misunderstood and mistreated outside the classroom environment. In this work, we consider individuals with development disabilities as individuals with cognitive challenges, autism, or Down syndrome. Previous research has shown that interactive technology offers several desirable advantages for such individuals, including (a) controllable input stimuli ([5],[18],[32],[33]), (b) multisensory and safer learning environment ([1],[2],[3],[37]), (c) opportunities for customization for individualized learning ([1],[2],[7],[18],[19]), (d) structured, predictable and consistent learning environment ([1],[2],[18],[32]), (e) the possibility to introduce controlled modifications or difficulty levels ([1],[2],[7],[18],[19]), (f) assistance in generalizations between scenarios ([1],[2],[31],[32],[33]), and (g) possibility for self-paced repetition of learning activities ([18],[19],[32],[33],[37]).

However, research on gesture-based applications for individuals with developmental disabilities is mainly focused on therapeutic interventions and specific social, motor and cognitive skills. The potential benefits of designing and developing real world task-based interactive applications that promote essential life skills for self-efficacy are largely under-researched. For the purposes of interventions, complex social and cognitive tasks can be broken down. For example, buying something from a local store requires social interaction - talking or asking for something, and cognitive skills, like knowing what to ask for, knowing how much money there is to spend, and deciding if the item and cost are comparable.

There are several options for implementing interactive applications. Although virtual reality applications are known to assist individuals with developmental disabilities in understanding real world challenges [6] and [23], it can be argued that for children on the autism spectrum the fidelity of the representation is less important [32], and visual, auditory and motor experiences and feedback, when used together, result in more efficient learning and information retention [6]. This notion is also supported by the theory of embodied cognition [42]. Thus, simpler 2D interfaces employing immersive gesture-based interaction can potentially be sufficient to simulate real world environments.

However, novel technologies are often perceived to not be cost-effective, as catering to a limited group of people, and difficult to maintain and integrate into existing systems. This is particularly true in developing regions, including India, where access to such technologies is still limited, especially for individuals with developmental disabilities. Moreover, available resources are low, integration and inclusion of individuals with development disabilities into mainstream society is strained, and the digital divide – technical and economic barriers – is more pronounced. To overcome these challenges and perceptions, it is important to substantiate the potential of such technologies to build a stronger case for their mainstream adoption. These applications also need to consider the cultural implications of the real environment, as “what constitutes appropriate behavior is largely a social construct” [3]. Thus, it is important to collaborate with the different stakeholders, for example, teachers, therapists and caregivers, and follow a user-centered design approach.
This paper presents the potential benefits of using a gesture-based application that simulates a real world scenario to impart a life skill to individuals with developmental disabilities. We first conducted a user-centered design study to identify a suitable gesture vocabulary and life-skill. We designed and developed an application for mimicking the real world scenario of buying groceries from a local mom-and-pop store, Kirana, in New Delhi. We evaluated Kirana with 18 individuals with developmental disabilities, who were unable to shop independently. Our findings suggest that applications that employ gesture-based interaction to simulate real world scenarios in a safe and controlled environment can provide learning that is translatable from the virtual to real world. Our main contribution is demonstrating the potential of gesture-based applications to facilitate self-efficacy in individuals with developmental disabilities.

In the following, we first contextualize the research with the related work in this domain. This is followed by a detailed description of our user-centered design study and its results. We then present the application description, evaluation methodology and its results. We conclude by discussing our findings.

2. RELATED WORK
In the related research we (a) discuss gesture-based interactions for individuals with developmental disabilities and its potential benefits vis-à-vis the theory of embodied cognition, (b) present an overview of assistive technologies for Indian children, and (c) differentiate our work from previous work on virtual reality applications for learnings real world skills.

2.1 Gesture-based interaction
Employing gesture-based, embodied interaction for social, therapeutic and educational applications for individuals with developmental disabilities has gained momentum in recent years. There is strong evidence in support of embodied learning paradigms based on the theory of embodied cognition, whereby cognition is situated within the environment and learning occurs also through bodily interaction with the environment [22], [42]. The embodied learning paradigm has been extensively studied in neuroscience and cognitive sciences ([10],[15],[21],[40]), and has been adopted by research in educational technologies within the human-computer interaction (HCI) domain [11].

Studies by Bartoli et al. [1],[2] showcase the benefits of embodied learning via gesture-based interaction for children with autism. Our work expands Bartoli’s work by extending it to individuals with developmental disabilities (including autism and Down syndrome), and to children from developing countries, by focusing on a life-skill. In line with Bartoli’s findings, we also purport that learning from gesture-based interaction is applicable to real world scenarios where selection via pointing and moving of physical objects simulates interactions in everyday life.

Previous research in this space is largely focused on individuals with autism and in improving social, motor and cognitive skills. The focus of the research has been on attention and memory or the concept of self, and it has employed tangible interaction and/or sensory motor perception. For example, SensoryPaint is multimodal application that incorporates tangible interaction and whole–body interaction for therapeutic interventions to encourage social interaction [35]. Research by the Lakeside Center for Autism1 and Kinems.com2, taps into the potential of gesture-based interaction and its inherent affordances for kinesthetic learning experiences within the classroom environment. MEDIATE [13] is a multisensory interactive environment that utilizes real-time visual, aural and vibrotactile stimuli. There are also studies on gesture-based applications to match visual facial expressions [7], improve hand-eye coordination, attention and focus [1], detect repetitive behavior or tantrums [13], for promoting joint attention [38], and for cognitive rehabilitation and exercising [14]. However, there is limited research on using gesture-based applications to impart life skills, such as purchasing items. In this context, incorporating culturally sensitive gestures and interactions is particularly important [3].

2.2 Assistive technologies in India
There is a culturally misguided attitude towards children with disabilities in India. A study of teachers’ attitude towards children with disabilities in schools in Mumbai showed that prior acquaintance with a person with a disability was a governing factor for teachers’ to be more positive and welcoming towards inclusive education [29]. A World Bank survey in 14 developing countries, including India, indicated a “worrisome vicious cycle of low schooling attainment and subsequent poverty among people with disabilities in developing countries” [12]. While there are several schools for children with special needs across India, very few of them employ technology within their classrooms and there is limited research examining the role of assistive technologies. One study proposed developing assistive communication technologies for individuals with autism or dyslexia in India [37]. Another, called Jollymate, is a digital notepad for children with dyslexia that emulates a “phonetics system of teaching letter sounds and letter formation” [19].

There are challenges in introducing new technological interventions, especially for individuals with development disabilities in the developing world: (a) resource constrains within the environment, for example, infrastructure and access to electricity, (b) a huge digital divide, thus communities that can benefit the most from technology have the least access to it, (c) inclusivity and integration among children with disabilities and typically developed is low, (d) technologies are too costly, especially for individualized use, and (e) stronger cultural barriers for individuals with disabilities that lead to more pronounced digital exclusion even within the technology-capable communities. For example, economically stable and educated parents might provide a mobile phone to a typically developed child and but not to a child with autism. In fact, studies have shown that cultural, societal and socio-economic factors largely affect “the experience of autism” [3] [16]. However, we believe that several emerging technologies are now affordable and one device can cater to a larger number of children, so that a whole school can time-share the resource.

2.3 Simulating real world scenarios
Simulations of real world scenarios to provide learning tools for individuals with special needs have been studied from the late 90s. This work has primarily focused on virtual reality environments (VR) utilizing computer generated three dimensional (3D) worlds. Applications include therapeutic, social and skill-based learning, including solutions for individuals with developmental disabilities, most notably autism and ADHD [8]. Studies have examined the benefits for social interactions [20] [18], collaborations with peers [31], or avatars [27], understanding facial expressions [17], pediatric rehabilitation [33], physical rehabilitation [23] [34], sense of presence [26], [31], [41], and skill based learnings [39]. For example, Coles et al. [6] and [28]...

---

1 http://lakesideautism.com/tag/kinect/
2 www.kinem.com
taught road and fire safety skills to children affected by prenatal alcohol exposure using. However, much of this research is focused in the developed world, and so the applications are possibly socially and culturally unsuitable for developing countries [3]. Our work does not utilize 3D virtual reality, but a simplified 2D graphical representation with gesture-based interaction. The graphics and gestures for interaction are designed from a real world scenario.

3. USER-CENTERED DESIGN STUDY

We worked with one of India’s oldest and most established special schools, called Nai Disha. Nai Disha is a part of Tamana.org, a nonprofit organization with four centers across Delhi. Nai Disha is dedicated towards providing young adults with developmental disabilities physical, emotional and functional independence by imparting vocational training and skill development. During the school’s annual Diwali Mela (fair celebrating the Indian festival of lights), we conducted a user-centered design study by setting up two games in one of the school’s classroom. The study had several purposes:

1. Capturing teachers’, students’ and their parents’ initial reactions to such a system
2. Defining a gesture vocabulary for three main interaction goals, namely navigation, selection and object manipulation
3. Identifying application areas or topics where gesture-based interaction can assist the students.

Setting up the installation during the Diwali Mela provided an event to introduce the researchers to various stakeholders - school staff, teachers, children and their parents - in an informal public gathering. This helped create an open and relaxed atmosphere to experience and discuss gesture-based applications, and an opportunity to gain insights into the stakeholders’ expectation and acceptance of such technology.

We designed two applications for the study to understand how the children interact with the screen space and select objects using free form gestures. The games provided an opportunity to identify gestures, which are fun, enjoyable and comfortable. The trials were recorded in short video clips, photos, questionnaires and observations. The data was analyzed by the researchers. The two applications and our findings from the user-centered design study are described next.

Figure 1: (left) a researcher demoing the painting game, (right) interface of the flash card based animal matching game.

3.1 Free-form painting

The free-form painting application was used as an introduction to gesture-based interaction without pre-defined tasks. The objective was to observe natural, comfortable and intuitive hand gestures, a novelty in the educational setting, and to derive gestures for navigating the screen space. Participants could use their hands as paint brushes and draw on a white canvas. The right hand would draw in red while the left in blue. The width of the hand-brush was controlled by moving the hand away (thinner) or towards (thicker) the white canvas. The participants were free to draw any shape or image they liked, as shown in Figure 1(left).

3.2 Animal matching game

The animal matching game was based on flash cards used extensively in Autism interventions, and for example also in the TOBY Playpad [43]. The game used pointing, with a dwell time of one second for selection, and drag-and-drop for object manipulation. There were three cards on the top row and three placeholders on the bottom row (Figure 1, right). The top row of cards was initially stacked and opened up after a swipe gesture was made, similar to the gesture of spreading out a deck of cards.

Participants selected a card by pointing at it using their right hand for one second. A successful selection attached the card to the cursor following the right hand. The attached card had to be dragged close to the correct placeholder in the row below. Thus, when a card (animal picture) was near its correct placeholder (text), the card animatedly flew on top of the placeholder, resembling magnetism [25]. To make the game simpler for students who could not spell, the flash cards could also be matched based on the background colors.

3.3 System description

Both of the applications were based on an in-house framework, utilizing the Microsoft Kinect sensor for gesture recognition, and consisting of three main processes; a Kinect service, graphics engine and application core logic. The Kinect service is a thin client over the Microsoft Kinect SDK that connects to the core logic. The Kinect is used to track the user’s body movements. For the two applications, we tracked only the upper body joints for the gestures. The graphical content is rendered using the Panda 3D\(^3\) engine. The core logic is a Python\(^4\) based application that takes the Kinect data as input, based on which the relevant content is displayed via the graphics engine.

Figure 2: Game Setup showing the interaction space

The system responds to the user closest to the device in front of it in a 1 meter by 1 meter area, 1.5 meters away from the Kinect (see Figure 2). We call this the active area, that is, the area in which a user can interact with the application using gestures. An onscreen hand cursor was present in both the applications to help guide the gesture interaction. Each applications starts with a verbal welcome message when a user enters the active area, followed by a short music clip and the screen fades out from black to the game screen. While the user is in the active area, the system responds to her gestures as defined by the game. When there is no longer a user in the active area, the game screen goes back to black.

\(^{3}\) https://www.panda3d.org/
\(^{4}\) https://www.python.org/
3.4 Procedure
The students visited the Diwali Mela setup with their parents and were encouraged to try both applications. Students interacted one by one, first with the free form painting, and then the animal matching game. Students were asked to ‘paint’ something with the first application and their parents usually suggested an object, such as a tree, or to write their name. Instructions were kept brief to identify gestures that are intuitive and natural. Students interacted for as long as they liked, usually stopping once the white canvas was colored over 70%. For the second application, students were asked to ‘match the animals’. The matching game was based on the already familiar flash cards, and students understood the task quite easily, and were able to finish the task.

Waiting students and parents, and other audience members were instructed to clap after every student interacted with an application. This was done as a reward for the students who tried out the unfamiliar environment and as a motivation for subsequent participants. Parents were interviewed after their child interacted with both of the applications and requested to fill in a feedback form. The feedback questionnaire aimed to understand the parents’ experience of watching their child try the applications and their own expectations and initial reactions to the technology.

3.5 Findings
In total, 18 students with developmental disabilities, that is, individuals with mental retardation, autism, or Down syndrome, (IQ M=49.78, SD=10; SCQ M=51.6, SD=10) [36] participated in our user-centered design study out of which parents of 14 of them filled in the responses after their child had interacted with both the applications. For the remaining four parents’ responses, either the parent was in a hurry or their child had interacted very briefly with only one of the applications.

3.5.1 Gestures for navigating the screen space
The onscreen cursor made it visually possible to imagine the hand-cursor as a brush to paint a red or blue line on the white canvas. From our observations of the gestures performed, the students formed circles using both hands with ease, as shown in Figure 3. Several participants, on the insistence of their parents, would try to write their name with large alphabetical gestures. Thus, we decided to design interaction-gestures using both hands with onscreen cursors for our main application.

3.5.2 Gesture for selection
There were two main gestures that were performed for selection: dwelling on an object or punching a fist towards the object. The punching gesture was popular and fun, but was not considered socially appropriate by the parents and teachers present. This is because the school experts shared concerns over the students’ tendency towards repetitive behavior and the socially negative impact of a punching gesture outside of gameplay. This finding resonates with the role of culture and context in defining socially acceptable gestures [3].

For selection via pointing with a dwell-time, we observed that the dwell time of one second was too short for the participants and it triggered several unintentional selections, as the several participants verbally exclaimed ‘oh’ when a card was attached to their onscreen cursor. Thus, for our main application, we decided to use the dwell time of 1.5 seconds for selection. A participant making a selection by pointing is shown in Figure 4(left). Our applications also worked well for a participant using a wheelchair without using the Microsoft Kinect SDK’s seated mode for gesture recognitions. Thus, with gestures requiring only upper body motion tracking, applications can be inclusive and comfortable for a larger group of participants.

3.5.3 Gesture for object manipulation
The animal matching game required moving a selected animal picture from the top row to a placeholder on the bottom row using the gesture inspired by the desktop drag and drop metaphor. Although all participants matched the three animal pictures to their placeholders, it was observed to be rather by chance than by choice, since the placeholders were magnetic. Thus, for object manipulation drag and drop was observed to be difficult. Figure 4 shows a participant being guided by her mother for the animal matching game. For our main application, we decided not to employ the drag and drop. Further research is required to identify a suitable gesture for this purpose.

3.5.4 Parent questionnaire
All of the parents agreed that their child enjoyed using the system and only 21% said it was tiring to use. Three out of the 14 children had used some form of gesture interaction before. Initial reaction of parents included excitement (50%) or being impressed (50%) by the games. 79% of the parents were enthusiastic about using gestures for educational interventions, while only one parent was skeptical and two did not comment. All but one parent had a positive overall impression of the system and that one parent had neither negative nor positive impression. 12 out of the 14 parents said they would consider using gestures for educational interventions in the future, while one parent wished the system worked better. Several parents, and also teachers, appreciated the inherent physical nature of the interaction, and commented that such games are good for body movement and encouraging physical activity. They hoped to see more rewards and motivation within the gameplay to encourage their child.

3.5.5 Potential application topics
Together with the teachers and parents at the Diwali Mela, we conducted a short focus group exercise to identify topics that are currently difficult to teach with traditional pedagogical methods and would greatly benefit from real world simulations with gesture-based interaction. These topics included monetary transactions, time management, planning or scheduling of events, working with constraints such as time or a budget, and the concept...
of currency (mainly because the size/shape/weight of Indian coins and notes does not relate to its value). Moreover, several teachers suggested designing an application that can simulate real world scenarios to assist in learning a life skill.

One of the life skills deemed important by the parents and teachers, was being able to purchase day to day items from a local kirana store, which are grocery stores in India (refer to Figure 5). The teachers explained that several students, who are able to communicate with strangers, are taken to the nearby market to practice buying an item, such as toothpaste, from the local store using a certain amount of money. However, sometimes a student is socially mistreated by strangers or the shopkeeper. Moreover, not all students are comfortable with all the processes involved in a visit to the local store, and the teachers are unable to extend the visit to these students. The school does not employ any other methods for teaching transactional calculations than traditional math classes and the occasional field trips to the local market.

Based on these discussions, it was decided that the real world scenario of buying an item from a Kirana store would be simulated in a gesture-based application and evaluated with the students who were not included in the monthly visits to the local store near the school. The details of the application, its interaction and design decisions are explained next.

![Figure 5: A kirana store in India](image)

4. KIRANA APPLICATION

In designing the Kirana application we broke down the life-skill into several smaller tasks: knowing the items to buy (decision making), asking for them from the shopkeeper (social interactions), looking up the price for each item, knowing if an item can be bought with the available cash (arithmetic), handing over the cash (social interaction), calculating the balance (arithmetic), and taking the balance and items (social interaction).

In India, each item has an M.R.P. (maximum retail price), which is required by law to be printed on the item cover, and is usually the price charged by stores in New Delhi for the item. Therefore, students are also taught the concept of M.R.P., which is contextually and culturally relevant for them.

These smaller tasks, which are independently achievable, bring out the bigger goal of ‘buying groceries’. Furthermore, the application aimed to promote socially acceptable and expected behaviors within the Indian context of Kirana shops, making the learning from the application, when translated to the real world environment, culturally appropriate. We did not include a shopkeeper-avatar in the application, as simulating social interactions to the required degree is complex. The application was designed to allow a teacher or moderator to imitate the social interactions based on individual needs.

Our research goal was to validate the potential of gesture-based interaction in translation of learning of a life skill from an application to real world scenarios for individuals with developmental disabilities, with a focus on the Indian social and cultural context.

The Kirana interface simulates a typical store layout. Customers stand outside the counter of the store and point to items they want to purchase. The screen has two shelves behind a table counter-top containing food items that can be bought, as shown in Figure 6. The items are randomly arranged at the start of the sessions, removing learnability of item placements. The application was built on the same in-house framework as the games in the user-centered design study applications described in section 3.3. The setup was also similar, that is, participants could interact within the 1 m² active area 1.5 meters from the Kinect (see Figure 2).

![Figure 6: Kirana- buying items on the table by paying using a 10 rupees note](image)

Based on the user-centered design study findings (section 3.4), gestures for interacting with the screen space and object selection used an onscreen cursor. Pointing with a dwell time of 1.5 second selects an object (item and money), which then animatedly slides to the table. This interaction eliminated the need of an explicit drag and drop gesture, which was found to be difficult (section 3.5.3). Moreover, since the application focus is primarily on three aspects of purchasing (decision making, social interaction and mathematics) the explicit gesture of handing over money was simplified in the application.

The session starts when a participant stands in front of the Kinect and is greeted by a female voice welcoming her to Kirana. The right side of the screen has the available money with the total amount displayed at the bottom. The left side shows the billing of items as they are bought allowing the participants to see the list of items they have bought and the total bill. The bill format is in line with Kirana shopkeepers who provide a written bill for all items purchased. The item’s price is automatically added to the bill on the left. Once there is at least one item to pay for, the participant can select any denomination of money using her right hand. This two-handed selection mechanism was added to encourage increased bodily movement and isolate the item selection from the money selection. Each item and money is accompanied by its spoken name and each transaction process has audio feedback. For example, when the balance is returned, the female voice says here is your balance. The application caters to the various endings: running out of money, or not enough money left to buy an item. For these scenarios, the female voice informs, for example, you have no money left. This is followed by a textual well done and an audio feedback to indicate the end of the session.

The animations of items and money were slow and sequential: there was only one item or money movement, from the shelves or
the wallet, to the table at a time. Once an animation was finished, the balance was returned back from the table to the wallet. Updating the bill and total amounts was also in sync with the series of animation. The application also supports customization; the type of item, its price and total available money in the wallet on the right hand side could be changed between sessions. However, this feature was not used in the evaluations.

5. EVALUATION AND RESULTS

The evaluation consisted of four phases. In the phases I and III, we conducted manual mathematical tests as a part of pre and post evaluation trials. These tests aimed to ascertain the mathematical ability (addition and subtraction) of the participant. The tests included single-digit subtractions, 2-4 digit additions, and single digit multiplications. The mathematical tests were evaluated by a teacher and a score was provided for each participant.

In phase II, Kirana was installed in classrooms and sessions were conducted for three weeks - with every participant once per week. The setup is shown in Figure 2. A line was marked 1.5 meters away from the Kinect to help the students’ position themselves for good gesture recognition. The participant was given a grocery list by the teacher and a fixed amount of money which she had to use to purchase the items from Kirana. Each participant was asked to complete the following tasks per session:

(i) You need to select items for yourself for breakfast from Kirana using your grocery list. For example bread, chips, milk and biscuits.
(ii) You have a budget of 100 Rupees to pay for the items you selected. Also check the balance returned by shopkeeper.

Data from the sessions included automated system logs with task times, items bought and monetary transaction details, moderator observations and a behavioral analysis of the participant to record positive and negative emotional behaviors or signals.

In phase IV, participants visited an actual kirana store near the school and were asked to buy several items. A moderator observed participant behavior with respect to the three sub-tasks: decision making (choosing an item from the list to buy), social interaction (talking to the shopkeeper and asking for an item) and mathematical ability (knowing how much to pay and balance to expect). Pre-shopping evaluations were not carried out because all participants were previously unable to shop in classrooms and sessions were conducted for three weeks - with every participant once per week. A moderator observed participant behavior with respect to the three sub-tasks: decision making (choosing an item from the list to buy), social interaction (talking to the shopkeeper and asking for an item) and mathematical ability (knowing how much to pay and balance to expect). Pre-shopping evaluations were not carried out because all participants were previously unable to shop according to their parents and teachers, and we did not want to make them uncomfortable. In order to understand the learning offered by the application, participants were not taught mathematical concepts in other class sessions during the evaluation.

The evaluations were conducted by two school teachers. This reduced any anxiety and complex social dynamics the participants might experience due to an unfamiliar presence. Teachers are also better equipped to understand nuances and implications of the participants’ behavior and reaction to the application or its tasks, responsible for and involved in all technical educational interventions within the school, and had participated in the design and development of the application. A workshop was conducted to discuss the evaluation goals and data collection requirements.

5.1 Participants

18 individuals (identified as P1-P18 in the following) with developmental disabilities participated in our evaluation (5 individuals with Down syndrome, 8 individuals with mental retardation, 3 individuals with autism and 2 individuals with cerebral palsy and mental retardation). There were 14 males and 4 females aged between 16 to 39 years ($M=26$, $SD=5.4$) with IQ ($M=46$, $SD=11$) and SCQ [36] ($M=58$, $SD=16$). The participants were recruited from the eighth grade of the school and they (a) understood the concept of left and right, (b) communicated verbally with the mediators, (c) understood instructions given to them, (d) had an awareness of self and body, (e) did not participate in the user-centered design study (mentioned in section 3), and (f) were previously unable to shop independently.

5.2 Results: Overview

While almost all the participants were excited and happy to play the game, two of them reported interaction fatigue. Four participants were observed to have limited left hand movement for two-hand interactions (items were selected using the left hand) although they did not explicitly mention any fatigue or pain. We also observed that the time taken to select items at the center of the screen was longer because five other participants first pointed to top-left corner of the screen and then moved towards the bottom right. Further research in required to identify optimum interaction gesture paths for individuals with varying motor capabilities.

The sessions were conducted during an Indian summer, and so surprisingly we observed that two participants expressed their desire to be able to buy cold beverages instead of food items from Kirana. One of the participants was upset at not being able to play more frequently because of her therapy schedules. We also observed that most of the participants felt bad if they were unable to complete a task or if they felt they were slow. At one instance, we noticed a participant slap her left hand to express her frustration.

With each progressive session, participants were more expressive and verbal than usual, and as expected, also showed improvements in the sub-tasks. For example, several participants navigated more carefully to items in the later trials. During the sessions, a line was drawn on the floor to enable participants to know the exact location from where to interact. In the later trials, we observed improvements in terms of understanding where to stand and reposition themselves.

5.3 Results: phase I and III

Figure 7 shows mathematical scores from phases I and III for each of the participants. In phase I (pre-trial) the highest score was 9.5 and lowest was 3 from a maximum 10 marks. In phase III, the lowest score was 5 and highest was 8. A score of 3 indicates knowledge of numbers from 1-10 and familiarity with money mainly notes, while a score of 6 indicates knowledge of numbers
from 1-100, familiarity with notes and coins, and simple two-digit additions and single-digit subtractions. A score of 10 would ideally mean that the participant has knowledge of numbers from 1-100, is familiar with notes and coins, and is comfortable with mental arithmetic including single-digit multiplication.

We observed an average improvement of 8.3% between phases I and III. Four participants showed a decrease in performance. It can also be noted that participants with a lower score benefitted more from the application than those with a score of 6 or above in phase I. Further research is required to observe whether long term evaluations would be beneficial for participants with a score of 6 or above, or whether they require increased difficulty in tasks.

5.4 Results: phase II
A total of 218 items were purchased by the participants in the three-week pilot. During the first week, the participants spent an average of 9.3 minutes (SD=4.5) per session buying while during the second and third week they spent 6 minutes (SD=4.4) and 4.5 minutes (SD=2.2) respectively. Consequently, participants spent an average of 2 minutes 34 seconds per item (SD=1.39) during the first week, 1 minute 43 seconds (SD=1.94) during the second week and 1 minute 13 seconds (SD=0.57) during the third week. This reduction in buying time per item indicates that participants became comfortable with the transactions and interactions with the application, over the course of the evaluations. We note here the limitation in measuring task time; we cannot differentiate the time taken for decision making and that of the actual interaction.

The participants spent an average of 71 Rupees during the first week, 80 Rupees during the second week and 63 Rupees during the third week. The moderators gave each participant an option to either select the items to purchase from the list or to select on their own. Surprisingly, we observed a steady increase in the preference to buy the items from the list (16% in session 1, 50% in session 2 and 73% in session 3) over the three sessions.

5.5 Results: phase IV
Actual store visits in phase IV provided highly ecologically valid way of translating the learning from virtual to real world scenarios. 12 of the 18 participants were taken to a local kirana store near the school in one session. Four participants (P3, P5, P16 and P18) were not a part of this activity because the teachers assessed they needed more practice with the application and two participants (P2 and P11) had moved to another school after phase III. It should be noted this was the first time the participants had come to an actual store by themselves. A moderator observed the participants interact with the shopkeeper and buy items from a given list – similar to the task in phase II.

5.5.1 Decision-making
Participants were able to connect the tasks during the visit to the tasks in Kirana, and purchase items from the given shopping list. The moderator observed a high degree of cooperation among the participants and they helped each other out wherever possible. This included, for example, crossing the road. The most popular items during the visit were cold beverages due to the Indian summer.

5.5.2 Social interaction
Participants who were feeling shy would observe the others interact with the shopkeeper before approaching the store (P9, P12, P15, and P17). One participant (P17) made the valid observation that everyone should take a bill from the shopkeeper so that the purchasing amounts can be checked by an adult later. Surprisingly, three usually nonverbal participants (P12, P13 and P14) were very responsive and able to interact with the shopkeeper.

5.5.3 Mathematical ability
Participants who had a high mathematical score (greater than 6) were comfortable with the concept of money and transaction involving arithmetic (P1, P4, P6 and P9). However, two participants gave all money with them to the shopkeeper: one out of nervousness (P12) and the other (P7) had problems reading the decimal place in price because of the small font size on the packet.

Overall, all the participants were able to locate an item’s M.R.P., understood that money is exchanged when purchasing the item, and expressed a positive attitude towards purchasing. However, as expected, the arithmetic aspect of calculating costs and balance was observed to be highly dependent on one’s mathematical ability. Several participants understood the relation between items and costs such that they only selected items they knew they would be able to purchase with the amount of money they had.

6. DISCUSSION
The following summarizes the benefits of our work as uncovered during our user-centered design process and Kirana evaluations. With our promising findings, we wish to encourage others researchers to work on applications that simulate real world environments to promote self-efficacy for individuals with developmental disabilities.

6.1 Promoting self-efficacy
In our evaluations, we observed improvements in mathematical ability, based on the teacher’s assessment, and translatable learning from the application to an actual local store. We believe that our application does the necessary groundwork in preparing individuals with developmental disabilities for purchasing items of need from local stores, thereby teaching them a valuable life skill that promotes self-efficacy.

However, when real life skills are being taught, particularly to individuals with developmental disabilities, cultural aspects must be considered. This is particularly relevant in the developing world where social awareness about individuals with disabilities varies greatly. The socio-cultural norms of a society dictate the degree of inclusivity and integration of individuals with developmental disabilities [3] and [16]. Additionally, individuals with developmental disabilities, such as autism, display a tendency towards repetitive behavior and teaching socially unacceptable gestures might increase their isolation, even if the gesture is fun during gameplay [1].

We developed an application that was culturally relevant and socially acceptable to the community by following a user-centered design approach and including the various stakeholders and decision makers – parents, caregivers, teachers, and therapists. Thus, applications that promote self-efficacy should employ gestures and interactions that are socially and culturally relevant and acceptable for the specific life-skill.

6.2 Simulating real world scenarios
Gesture-based applications can simulate real world scenarios in a safe and controlled environment through interactions that are appropriate and relevant to a life-skill. Using free form body gestures provides a mechanism for more inclusive social interaction and team work, especially when compared with touch screens that have limited form factor and surface area to support multiple users. The inherent nature of embodied interaction [42] makes for an immersive and engaging experience. In Kirana, even...
though it only supported a single active user, the interaction was visible to the moderators who could then prompt and encourage the participant, thus also involving them in the scenario.

6.3 Designing for technology acceptance

We note here the collaborative nature of our work, from the user-centered design study to the Kirana evaluations and analysis. Since the application was designed and developed with teachers, therapists and parents, who provided valuable insights and experience, there was no resistance towards technology acceptance or adoption. Thus, the technology was integrated within the classroom environment with ease.

Additionally, we identified ways to address the challenges in introducing new technological interventions for individuals with development disabilities in the developing world. First, economic barriers for technologies that are too costly can be overcome by designing applications that can be integrated within schools, and can be personalized and customized for use by a larger group of individuals. Second, resource constraints can be overcome through collaborations between schools, universities and industry partners. Third, the digital divide can be reduced by spreading awareness of the benefits of technology within schools and to parents.

Our main contribution is not the technology behind the Kirana application as such, but rather the focus on the currently overlooked potential of interactive technology for providing translatable learnings to real world scenarios, when using contextually appropriate gesture-based applications to teach life skills. In India, this technology is rare and novel, and thus its acceptance remains largely unexplored. Based on our findings, we can assume that following a user centered design approach helps reduce the challenges in technology integration.

6.4 Providing control to the teacher

Due to the wide range of abilities of an individual, the teacher or moderator should be able to customize the application based on individual capabilities and interest, as also stated in previous research by Bartoli et al. [1], [2], [38]. In Kirana, the teacher guided the social interaction and the task, for example, by giving a shopping list. This enabled the teacher to customize the learning, even on the fly. Moreover, number, type and price of items, and the total amount of money in the wallet could be changed easily between sessions. By providing control to the teachers, we believe we also increased the acceptance of technology within the classroom environment.

6.5 Providing multimodal feedback

Individuals with developmental difficulties find it easier to select items that have both visual (change in size) and auditory (name of the item is said aloud) feedback, instead of only one. This reinforced multimodal feedback helps overcome visual or auditory impairments, if any, and provides multiple stimuli for attention [1] [2]. In Kirana, each visual object, upon being selected, also had an auditory feedback.

6.6 Providing clear start and end of gameplay

To avoid ambiguity and confusion during interaction, it is extremely important to provide a clear start and end of gameplay. In Kirana, when the participant moves out of the Kinect’s active area, the application screen turns black to indicate an end of interaction. The application also catered to various end-scenarios such as not enough money left to buy an item, or running out of money or items to buy. Each of these scenarios stated the reason for the end, followed by a visual and audio “well done” feedback to conclude the session.

6.7 Providing serial and structured content

Overall, our findings show that gesture-based applications that simulate real world scenarios can be used to teach a life skill to individuals with developmental disabilities. This is achieved by designing interaction that is socially and culturally appropriate and breaking down complex tasks that require social, mathematical and decision-making skills. We recommend such complex tasks to be broken down into a sequence of smaller achievable steps. If the animation and other visual media content also follow the same step by step order, individuals with developmental disabilities can follow the progress of the tasks with ease. As an example, in Kirana, the purchasing animations were slow and sequential.

7. CONCLUSION

This paper presented an application that employs gesture-based interaction for teaching a real world skill, of buying groceries, to individuals with developmental disabilities in India. By following a user-centered approach and including various stakeholders and decision makers – parents, caregivers, teachers, and therapists – we developed an application that was culturally relevant and socially acceptable by the community. Kirana simulated the real world scenario of pointing to an item to buy it, while interacting with a shopkeeper at a local grocery store in New Delhi. The results of our evaluations show promising translations of learnings from the application to a real world context. Our findings provide strong support in favor of gesture-based systems for enabling teachers and educationists across the globe to impart life skills to individuals with developmental disabilities.

8. ACKNOWLEDGMENTS

Our thanks to the school for their valuable time and guidance, and a special thanks to all the participants. We also like to thank Sanna Grönlund for developing the graphics for the game.

9. REFERENCES


**Kirana: A Gesture-based Market App for Life Skills Learning for Individuals with Developmental Disabilities**

Sumita Sharma¹, Saurabh Srivastava², Krishnaveni Achary³, Blessin Varkey³, Tomi Heimonen⁴, Jaakko Hakulinen¹, Markku Turunen¹, and Nitendra Rajput⁵

¹University of Tampere, Tampere, Finland
²Xerox Research Centre India, Bangalore, India
³Autism Research Center, Tamana School of Hope, New Delhi, India
⁴University of Wisconsin-Stevens Point, Stevens Point, WI, USA
⁵IBM Research, New Delhi, India

{firstname.lastname, sumita.s.sharma}@sis.uta.fi, saurabh.srivastava@xerox.com, {krishnaveni.achary, blessinvarkey}@gmail.com, theimone@uwsp.edu, rnitendra@in.ibm.com

**Abstract**

*Kirana* is a gesture-based application that simulates purchasing experience at a local Indian grocery store. It provides individuals with developmental difficulties a safe and controlled environment to explore the processes of purchasing an item: namely, deciding items to buy (decision making), calculating costs and balance (mathematical skills), and interactions via pointing (social interaction). Previous research has established that gesture-based interaction has the potential to enhance social, motor and cognitive skills for individuals with developmental disabilities. Currently, translating learnings from gesture-based virtual applications to real world scenarios is underexplored. Kirana is an approach to simulating practical real interactions, by breaking down complex tasks that require social, mathematical and decision-making skills, and encouraging self-efficacy. Skills learnt by purchasing items with *Kirana*, are potentially transferable to a real world store.

**Keywords**

Gesture-based interaction; Developing countries; Individuals with developmental disabilities

1. **Introduction**

Individuals with developmental disabilities face several challenges in learning skills that promote self-efficacy. There is a threat of being mistreated and misunderstood outside of their home or school and difficulties in arranging multiple self-paced practice sessions in the real environment. When attempting to perform daily living tasks such as purchasing items at a store, individuals with developmental disabilities can be overwhelmed by the number of smaller subtasks involved – of deciding which items to buy, of calculating costs and balances, and interacting socially with the shopkeeper. Therefore, it is important to provide a safe, controlled and self-paced learning environment to practice buying of items before exposing the individual to the real world environment.

Working with special educators, parents and teachers from the Nai Disha Tamana School in New Delhi, we designed and evaluated an interactive gesture-based application, called *Kirana*. The application aims to teach the real life skill of purchasing items from local stores (“kirana”) in India. These local stores usually consist of shelves full of items along the walls and a large rectangular table at the store entrance. On top of the table is a cash register, among other items. These stores are usually quite small where customers stand outside the store and point to the items they want. The application interface and interaction simulated the real world scenario of buying items from an actual store, and they are described in the next section.

2. **Kirana Application**

The application’s interaction and interface was designed to break down the task of buying groceries into smaller achievable tasks. Moreover, by designing within the Indian context and with Indian users, we aimed to promote socially acceptable gestures and behaviors. This helps ensure any insights gained from the application are culturally appropriate when translated to a real world setting [3]. With Kirana, a teacher or care giver can assign a customized task, for example, buying items from a given list, or buying items for breakfast (with a given budget). The task is then completed by buying the items using the application. The application also provides easy to configure task elements - *types of items, price of items and total available money* - to allow for individualized learning goals. Moreover, tasks can be made increasingly complex to encourage users to sharpen their skills. *Kirana* was evaluated with participants from Nai Disha Tamana School in New Delhi which is described in detail in [3]. We next describe the interface and interaction of the application with respect to the real world scenario of purchasing an item from such a store.

2.1 **Interface**

The interface of the application is shown in Figure 1. It contains two shelves with various items on each. The number of items are limited to four per shelf to reduce the complexity of the overall learning task. In the actual store, there would be several shelves alongside the walls with enough items on each shelf to hide the wall behind it. The application also has a large rectangular wooden table in front of it which also resembles the tables or counter in actual stores. This table usually has a cash register on it, and that is why the items being bought and money being paid are also kept on the table. The right side of the interface shows a user’s wallet with the amount of money and the left side shows the current bill as one would expect from an actual store. We intentionally do not have an avatar for the shopkeeper for two reasons: to allow the mediator to incorporate personalized social interactions for each individual user, and to eliminate the complexity in attempting to simulate complex social interactions.
2.2 Interaction
The interaction with the application is kept as close to the real world scenario as possible. To select an item on a shelf, users have to point at it for 1.5 seconds using their left hand. Likewise, money is selected by pointing at it for 1.5 seconds with the right hand. This two-handed selection mechanism isolated the item selection from the money selection and increased bodily movements (moving both arms instead of one), which was deemed desirable by the teachers. Once an item or money is selected, it animately flies onto the table. We intentionally animated item movements to eliminate the drag and drop gestures from the application because it was found to be confusing and difficult for the target users [3]. In the real world scenario, the shopkeeper would usually move the items from the shelf to the table.

2.3 System Description
The system consists of a laptop running the application software and connected to an HD display and a Microsoft Kinect 360 sensor. The application was built using an in-house application framework (described in more detail in [2] and [3]), consisting of three main processes: a Kinect service, graphics engine, and application core logic. The Kinect service is a thin client over the Microsoft Kinect SDK that connects to the core logic over a TCP socket. Microsoft Kinect is used to track a user’s hand with respect to their body within a 3D pointing area, called the physical interaction zone (PhIZ) [1]. The zone is relative to the user’s body, and not the display. All audio and graphical content is rendered using the Panda 3D graphics engine. The core logic is a Python based application that consists of an input/output management module, handling all the inter-process communication and acting as an interface between the core logic, the Kinect, and the graphics engine (ibid).

2.4 Transaction Flow
A typical session with the application consists of the following:

- **Start**: A user stands in front of the system. Upon detection by the Kinect, the initial black screen fades out to the application screen shown in Figure 1. A female voice (pre-recorded) greets the user by saying ‘welcome to the virtual market’, which is also shown textually on the screen. To remove learnability of item placements, the items are randomly arranged each time.

- **Select items**: The user can now select items by pointing at them for 1.5 seconds using their left hand. Once an item is selected, it animately flies to top-left of the table, and the bill is updated with the item name and its cost. When browsing over items for selection, the item name is spoken out by the female voice.

The user can continue to select several items from the shelves until available funds are exhausted. That is, if the bill total will exceed the total money available, that item will not be selected and a female voice will remind the user that there is ‘not enough money left’. Thus, in this way, the application provides feedback for incorrect actions. However, the application intentionally does not provide additional prompts or hints to the user to allow them to spend as much time as they require on the task.

- **Paying for the items**: the user can select money by pointing at it for 1.5 seconds using their right hand. Money selected also animately flies but to the center of the table. Similar to the item announcement during selection, money denominations are also spoken by the female voice when browsing over them for selection. Money that is paid (selected) is reduced from the total bill amount to indicate the remaining cost. Once the cost is paid, the items fly into the blue shopping bag and any remaining monetary balance is returned to the table from where it animately flies back to the right side of the interface, and is added to the available money.

- **End**: A session ends when there is not enough money left to buy any more items or when the user walks away from the system (outside the tracking area of the Kinect sensor). This would typically take place after completing the desired transactions but can also be the result of the user accidentally leaving the area.

At the end of each session there is a loud applause and the female voice says ‘well done’ to all users.

3. ACKNOWLEDGMENTS
We thank all the Nai Disha Tamana School students for their time and efforts towards designing and evaluating Kirana. We are also grateful to Sanna Grönlund for designing the application interface.

4. References

---

1 developer.microsoft.com/en-us/windows/kinect
2 www.panda3d.org

Designing Gesture-Based Applications for Individuals with Developmental Disabilities: Guidelines from User Studies in India

SUMITA SHARMA, University of Tampere
BLESSIN VARKEY and KRISHNAVENI ACHARY, Tamana
JAAKKO HAKULINEN and MARKKU TURUNEN, University of Tampere
TOMI HEIMONEN, University of Wisconsin–Stevens Point
SAURABH SRIVASTAVA, Conduent Labs India
NITENDRA RAJPUT, InfoEdge India Ltd.

Gesture interaction provides a multitude of benefits to individuals with developmental disabilities, from enhancing social, motor and cognitive skills to providing a safe and controlled environment for simulating real-world scenarios. As gesture-based applications gain ground in the special education domain, we study their potential in the Indian context. Together with Tamana, an NGO in New Delhi, we have been conducting a series of exploratory user studies since October 2013. This includes the design and evaluation of three gesture-based applications to impart social and life skills to individuals with developmental disabilities. The Kirana application employs socially appropriate gestures to teach the life skill of buying day-to-day items from a local Indian grocery. Balloons promotes joint attention skills through collaborative interaction. HOPE improves motor coordination and social and cognitive skills, with increasing levels of difficulty. Based on studies with these applications, this article presents guidelines for designing gesture-based applications for individuals with developmental disabilities. The guidelines focus on (a) designing applications that cater to a larger group of individuals to encourage collaboration and inclusion, for instance, providing easy and controllable transitions between different task levels, and balancing interaction and content complexity; (b) addressing the challenges in conducting research in this domain, with respect to ethical and procedural decisions; and (c) designing for technology acceptance within the Indian context, for example, by following a collaborative and stakeholder inclusive approach, and addressing apprehensions towards technology adoption. These guidelines aim to benefit other practitioners working in this domain and especially in the educational technology context of India.

CCS Concepts: • Human-centered computing → Gestural input; Empirical studies in interaction design; Accessibility technologies;

Additional Key Words and Phrases: Gesture interaction design, design for all, HCI, individuals with developmental disabilities

Author’s addresses: S. Sharma, J. S. Hakulinen, and M. Turunen, Faculty of Communication Sciences, University of Tampere, Kalevantie 4, Tampere 33100, Finland; emails: {sumita.s.sharma, jaakko.hakulinen, markku.turunen}@sis.uta.fi; B. Varkey and K. Achary, Tamana School of Hope, CPWD Complex, Vasant Vihar, New Delhi 110057, India; emails: {blessinvarkey, krishnaveni.achary}@gmail.com; T. Heimonen, University of Wisconsin–Stevens Point, 2100 Main St, Stevens Point, WI 54481, USA; email: tomi.heimonen@uwsp.edu; S. Srivastava, Conduent Labs, Etamin Block Wing-A, 4th Floor, Prestige Technology Park-II, Marathahalli - Sarjapur Outer Ring Road, Bengaluru, Karnataka 560103, India; email: Saurabh.Srivastava@conduent.com; N. Rajput, InfoEdge India Ltd, B-8, Sector-132, Noida 201301, India; email: nitendra@acm.org.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2018 ACM 1936-7228/2018/03-ART3 $15.00
https://doi.org/10.1145/3161710

ACM Transactions on Accessible Computing, Vol. 11, No. 1, Article 3. Publication date: March 2018.
1 INTRODUCTION

Interactive technology offers several advantages and possibilities for individuals with developmental disabilities, including (a) a controllable input stimuli [5, 16, 28, 29]; (b) a multisensory and safe learning environment [2, 4, 32]; (c) customization for individualized learning goals [2, 3, 7, 16, 17]; (d) structured, predictable, and consistent learning environments [2, 3, 16, 28]; (e) controlled difficulty levels [2, 3, 7, 16, 17]; (f) assistance in generalizing to other similar scenarios [2, 3, 27, 29]; and (g) self-paced repetition of learning activities [16, 17, 28, 29, 32]. Moreover, applications can simulate real-world scenarios in a safe and controlled environment, and the learning is potentially translatable from the virtual to the real world, facilitating self-efficacy [8, 23, 30, 34, 36].

This article presents guidelines for designing and developing applications employing mid-air gestures for individuals with developmental disabilities. We consider individuals with developmental disabilities as individuals with cognitive challenges, including Down syndrome, attention-deficit hyperactivity disorder (ADHD), and high-functioning autism. We address the potential of designing for users in India, where there are several challenges towards adoption of novel technologies, from being expensive and catering to a limited group of people to difficulties with technology integration and maintenance. Furthermore, individuals with developmental disabilities have limited access to these technologies, and the digital divide—technical and economic barriers towards technology access—is more pronounced. Thus, it is important to substantiate the potential of such technologies and build a strong case for their adoption. Collaborating with different stakeholders, for example, educators, therapists, and caregivers, and following user-centered design approach can potentially increase technology acceptance. Moreover, socially acceptable interactions should be designed by taking into account the cultural implications of the environment [4].

The guidelines for addressing the above issues are based on our work on gesture-based applications with a nonprofit organization, called Tamana, at their Nai Disha School since October 2013. Tamana is dedicated to providing young adults with developmental disabilities physical, emotional, and functional independence by imparting vocational training and skill development. Our collaboration with Tamana has produced several educational applications. In this article, we present the exploratory user studies of three applications: Kirana teaches life skills by breaking down complex tasks that require social, mathematical, and decision-making skills; Balloons promotes joint attention skills through social collaboration (see Figure 1); and HOPE improves motor coordination and social and cognitive skills. This article is an extension of our prior work on Kirana, which was presented at the ASSETS 2016 conference [34]. In this article, we present results from the two aforementioned additional studies and summarize the findings from all the studies as a set of guidelines. The main contributions of this article are as follows:

Guidelines for gesture-based applications: We combine our results and experiences from user studies with Kirana, Balloons, and HOPE to present guidelines for designing gesture-based applications for individuals with developmental disabilities, with a focus on three aspects: designing the application, designing for research in this domain, and designing for the Indian context.

Redefining inclusivity in interactive technologies: The Kirana application [34, 35] and our previous study with Balloons [33] focused on a small group of individuals. With HOPE, we designed
Designing Gesture-Based Applications for Individuals with Developmental Disabilities

Fig. 1. (left) A participant using Balloons, and (right) a moderator assisting a participant with the pointing gesture (pictures adapted from Ref. [33]).

for inclusion-within, that is, an application that not only caters to a larger group of individuals but also encourages collaboration. This is motivated by Nai Disha School’s conscious efforts towards inclusion of students, within their schools and mainstream society. Therefore, user studies with Balloons and HOPE presented in this article include a larger spectrum of individuals with development disabilities.

In this article, first an overview of the studies is presented (Section 1.1) followed by related work in this domain (Section 2). Then, the applications and their evaluations are described starting with Kirana (Section 3), Balloons (Section 4), and HOPE (Section 5). Finally, combining the findings and experiences from all the evaluations, a collated set of guidelines for practitioners in this field is provided (Section 6).

1.1 Research Overview

The findings presented in this article are derived from the design journeys, exploratory user studies, and the experiences of the school specialists involved in the work with three applications:

Kirana: The Kirana application [34, 35] was designed to teach the life skill of buying items from a local Indian grocery store by breaking down social, mathematical, and decision-making skills into smaller subtasks. It employs pointing with dwell time for selection of items and eliminates the need of object manipulation gestures by animating item movements between the different subtasks.

Balloons: The Balloons application was designed for children with medium-low functioning autism [33] to promote social interaction via shared experiences. It employs pointing with dwell time for selection, and two users can interact with the application simultaneously. In this work, Balloons was used to introduce the participants to gesture-based interaction and the concept of colors.

HOPE: HOPE focuses on improving motor coordination and social and cognitive skills through a series of spatial reasoning tasks, employing a gesture sequence of grab, drag, and drop. In this work, the tasks are focused on matching colors, building on the concepts in Balloons, while introducing a complex sequence of gestures for selection and object manipulation.

Nai Disha School’s psychologists evaluate each student annually, or biyearly, and maintain a record of their developmental progress, which includes IQ and SQ (social quotient) scores. IQ is measured using an Indian adaptation of Wechsler’s intelligence scale for children [41], called Malin’s Intelligence test [21]. SQ is measured using the Vineland social maturity scale, which is a psychometric assessment designed to measure the social competence of an individual. It measures...
Table 1. Participant Information for All the Studies

<table>
<thead>
<tr>
<th>Participant Information</th>
<th>Kirana</th>
<th>Balloons and HOPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Participants</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>Age (in years)</td>
<td>M = 26, SD = 5.4</td>
<td>M = 29, SD = 7</td>
</tr>
<tr>
<td></td>
<td>M = 22, SD = 4</td>
<td>M = 22, SD = 4</td>
</tr>
<tr>
<td>Gender (Male, Female)</td>
<td>M = 14, F = 4</td>
<td>F = 9</td>
</tr>
<tr>
<td></td>
<td>M = 4, F = 1</td>
<td>M = 4, F = 1</td>
</tr>
<tr>
<td>IQ</td>
<td>M = 46, SD = 11</td>
<td>M = 42, SD = 11</td>
</tr>
<tr>
<td></td>
<td>M = 63, SD = 20</td>
<td>M = 57, SD = 15</td>
</tr>
<tr>
<td>SQ</td>
<td>M = 58, SD = 16</td>
<td>M = 41, SD = 10</td>
</tr>
</tbody>
</table>

eight categories of behavior: “self-help general, self-help eating, self-help dressing, locomotion, occupation, communication, self-direction, and socialization” [10]. Our work with Kirana focused on a narrow group of individuals towards the severe end of the developmental disability spectrum (IQ < 50, SQ < 60). In studies with Balloons and HOPE, participants were divided into two groups based on their IQ and SQ, whereby group one participants are on the severe end of the developmental disability spectrum (IQ and SQ < 50) and group two participants are towards the moderate-mild end of the spectrum (IQ and SQ > 55). A total of 32 unique participants are a part of the work presented in this article: Eighteen participants in the Kirana evaluation and another 14 in Balloons and HOPE. A summary of participant information for each study is presented in Table 1.

The guidelines presented in this article are therefore based on several different variations of gesture design, application content, and participant profile, providing a well-rounded perspective towards designing gesture-based applications for individuals with varying levels of developmental disabilities.

2 RELATED WORK

In the following, we discuss gesture-based interaction for individuals with developmental disabilities and its benefits vis-à-vis the theory of embodied cognition and present an overview of assistive technologies for Indian children.

2.1 Gesture Interaction for Individuals with Developmental Disabilities

Employing gesture-based, embodied interaction for social, therapeutic, and educational applications for individuals with developmental disabilities has gained momentum in recent years. Such embodied learning paradigms are based on the theory of embodied cognition, where cognition is situated within the environment and learning partially occurs through bodily interaction with the environment [19, 38]). The embodied learning paradigm is extensively studied in neuroscience and cognitive sciences [9, 14, 18, 37] and has been adopted by research in educational technologies within the human–computer interaction (HCI) domain [11].

There is noteworthy research on the benefits of gesture-based interaction for individuals with developmental disabilities. For instance, the Lakeside Center for Autism1 and Kinems.com2 have developed applications to improve social, motor, and cognitive skills for children with autism. With affordable motion-tracking technologies, such as with the Microsoft Kinect, several studies have also incorporated gesture-based applications to match visual facial expressions [7]; to improve hand-eye coordination, attention, and focus [1]; to detect repetitive behavior or tantrums [13]; and for cognitive rehabilitation and exercising [36].

1http://lakesideautism.com/tag/kinect/.
MEDIATE [25, 26] is an interactive multisensory environment that uses visual, aural, and vibrotactile stimuli in real time to motivate and engage children with autism. Children are allowed to explore the spaces at their own pace and there is no time limit to the interaction. Results from user studies in multiple European cities have shown a high acceptance of its interactive space, as it allows creative self-expression using multimodal applications, creating therapeutic learning experiences. Likewise, SensoryPaint, a multimodal application that incorporates tangible and whole-body interactions, provides therapeutic interventions to encourage social interaction among children with autism [31].

Furthermore, studies by Bartoli et al. [2, 3] showcase the benefits of embodied learning via gesture-based interaction for children with autism. In line with Bartoli’s findings, we also purport that learning from gesture-based interaction is applicable to real-world scenarios where selection via pointing and moving physical objects simulates interactions in everyday life. Bartoli et al. [3] also provide design guidelines for gesture-based playful interactions based on their ongoing research with children with autism in Italy. They divide their guidelines as general or goal specific, focusing on motor, cognitive, and social skills, which are important considerations for children classified with medium-low functioning autism. Our work expands Bartoli’s work by extending it to individuals with developmental disabilities (including autism, Down syndrome, and ADHD) and to usage in the context of developing countries.

2.2 Assistive Technologies in India

In India, there is a culturally misguided attitude towards individuals with disabilities. For instance, in a study of teachers’ attitude towards children with disabilities in schools in Mumbai, teachers were more positive and welcoming towards inclusive education if they had prior acquaintance with a person with a disability [24]. Moreover, there is “worrisome vicious cycle of low schooling attainment and subsequent poverty among people with disabilities in developing countries,” as reported by the World Bank based on a survey conducted in 14 developing countries [12]. There is limited research on the benefits of assistive technology in education, even though there are several schools in India that employ technology for children with special needs. One study proposed developing assistive communication technologies for individuals with autism or dyslexia in India [32]. Another, called Jollymate, is a digital notepad for children with dyslexia that emulates a “phonetics system of teaching letter sounds and letter formation” [17]. Tamana has incorporated TOBY Playpad [39], an iPad-based learning application, into their educational intervention for children with autism [40].

There are also several challenges in adopting new education technology, especially for individuals with developmental disabilities in the developing world. For instance, there can be limited access to resources within the environment, like infrastructure and electricity. There is also a growing digital divide where entire communities do not have access to potentially beneficial technologies. In schools, inclusivity and integration between individuals with disabilities and typically developed is low. Moreover, technology is considered too expensive, especially for individual use. There exist strong cultural barriers for individuals with disabilities, which increases digital exclusion even within technology-capable communities. In fact, studies have shown that cultural, societal, and socio-economic factors largely affect “the experience of autism” [4, 15]. For instance, economically stable and educated parents might provide a mobile phone to a typically developed child but not to a child with autism. However, costs issues are mitigated when emerging technologies become affordable, one device is customized to cater to a larger number of users, and resources are shared. With these changing attitudes, we believe our work to be timely and well situated within the Indian context. Next, we present the design and evaluation of the Kirana, Balloons, and HOPE applications.
3 KIRANA APPLICATION

To inform the design of Kirana, we conducted several discussions with different stakeholders at Nai Disha School’s annual fair. The fair provided an informal event for researchers to interact with various stakeholders—school staff, educators, students, and their parents—to gain insights into the stakeholders’ expectations and the challenges towards acceptance of technology. For the complete description of the annual fair’s exploratory design work, please refer to our previous work [34]. From these discussions, we identified one potential application area for gesture-based interaction: simulating real-world scenarios. In this case, the focus was on teaching how to purchase day-to-day items from a local grocery store in Delhi, called Kirana (shown in Figure 2). At Nai Disha School, students who are able to communicate with strangers have monthly visits to a local store to practice buying day-to-day items, such as toothpaste. However, sometimes a student is socially mistreated by strangers or the shopkeeper. Moreover, not all students are comfortable with the activity, and the school specialists are unable to extend these visit to those students. Our research goal was to validate the potential of gesture-based interaction in the translation of learning of a life skill from an application, practiced in a safe environment, to real-world scenarios.

3.1 Design of Kirana

In designing Kirana [35], we broke down the task of purchasing an item into several smaller sub-tasks: selecting the item to buy (decision making), asking for it from the shopkeeper (social interaction), looking up the price for the item, knowing if the item can be bought with the available cash (arithmetic), handing over the cash (social interaction), calculating the balance (arithmetic), and taking the balance and item (social interaction). In India, all products have a maximum retail price (M.R.P.), which is required by law to be printed on the packing and is usually the price charged by stores in New Delhi. Therefore, the concept of M.R.P. needed to be incorporated into the application to make it contextually and culturally relevant. Kirana’s interface simulates a typical store layout (Figure 2), where customers stand outside the store and point to items they want to purchase.

The application screen has two shelves containing food items that can be bought behind a table counter-top, as shown in Figure 3. The right side of the screen has the available money with the total amount displayed at the bottom. The left side shows the billing of items as they are selected. The bill format is in line with Kirana shopkeepers, who provide a written bill for all items purchased. The application has an onscreen hand cursor for interaction and the gesture for item selection is pointing with a dwell time of 1.5s. When an item is selected, its price is automatically added to the bill on the left. Once there is at least one item to pay for, the participant can select any
denomination of money using his or her right hand. This two-handed selection mechanism was added to encourage bodily movement and isolate the item selection from the money selection.

To simplify the interaction, given the complexity of the task, selected items automatically slide to the desired onscreen location, eliminating the need of an explicit drag and drop. The animations of items and money are slow and sequential: There is only one item or money movement at a time, from the shelves or the wallet, to the table. Once these animations are finished, the balance is returned from the table to the wallet, and the bill and total amounts are updated.

Each item and money are accompanied by its spoken name, and each transaction process has audio feedback. For example, when the balance is returned, a female voice says *here is your balance*. The application also caters to the various endings: running out of money or not enough money left to buy an item. For these scenarios, the female voice informs, for example, *you have no money left*. This is followed by a textual *well done* and audio feedback to indicate the end of the session.

### 3.2 Application Setup

The application is based on an in-house framework, utilizing the Microsoft Kinect sensor for gesture recognition. The Kinect tracks a user’s hand with respect to their body within a three-dimensional area, called the *physical interaction zone* or PhIZ [22]. PhIZ is relative to the user, and not the display, making it more flexible to point at larger distances (Figure 4, left). The system responds to the user closest to the Kinect, in a 1m by 1m area, 1.5m away from the Kinect (Figure 4, right). We call this the *active area*, where users can interact with the application using gestures. An
onscreen hand cursor helps guide the gesture interaction. A session starts with a verbal welcome message when a user enters the active area, followed by a short music clip, and the screen fades out from black to the game screen. While the user is in the active area, the system responds to their gestures as defined by the game. When the user leaves the active area, the game screen goes back to black. A large display was used with a laptop running the application.

3.3 **Kirana Participants**

Eighteen individuals with developmental disabilities participated in our evaluations (5 individuals with Down syndrome, 3 individuals with autism, 2 individuals with cerebral palsy and an unspecified form of intellectual disability, and 8 individuals with other unspecified forms of intellectual disabilities). There were 14 males and 4 females aged between 16 and 39 years ($M = 26$, $SD = 5.4$) with a mean IQ of 46 ($SD = 11$) and SQ of 58 ($SD = 16$). The participants were recruited from the eighth grade of the school and they (a) understood the concept of left and right, (b) communicated verbally with the moderators, (c) understood instructions given to them, (d) had an awareness of self and body, and (e) were previously unable to shop independently.

3.4 **Kirana Evaluation Procedure**

The evaluation consisted of four phases, starting in April 2014, and was moderated by two special educators. This reduced any anxiety and complex social dynamics the participant might experience with an unfamiliar researcher. The educators were also better equipped to understand the nuances and implications of the participants’ behavior and reactions to the application or its tasks. They were also responsible for and involved in all technical educational interventions within the school and had participated in the design and development of the application.

*In phases I and III*, manual mathematical tests were conducted for pre- and post-evaluation. These tests aimed to ascertain the mathematical ability (addition and subtraction) of the participant and included single-digit subtraction, two- to four-digit addition, and single-digit multiplication. The tests were evaluated by the educators and a score of 10 was provided for each participant. To understand the learning offered by the application, participants were not taught mathematical concepts in other class sessions during the evaluation.

*In phase II*, Kirana was installed in a classroom, and sessions were conducted for three weeks, with every participant interacting once per week. The participant was given a grocery list by the educator and a fixed amount of money from which they had to use to purchase the items. Each participant was asked to complete the following tasks per session:

- You need to select items for yourself for breakfast from Kirana using your grocery list. For example, bread, chips, milk, and biscuits.
- You have a budget of 100 rupees to pay for the items you selected. Also check the balance returned by the shopkeeper.

*In phase IV*, participants visited an actual Kirana store near the school and were asked to buy several items. One of the educators observed participant behavior with respect to the three sub-tasks: decision making (choosing an item from the list to buy), social interaction (talking to the shopkeeper and asking for an item), and mathematical ability (knowing how much to pay and balance to expect).

*Data Gathered:* Data from the sessions included automated system logs with task times, items bought, and monetary transaction details, moderator observations, and a behavioral analysis to record positive and negative emotional behaviors or signals.
3.5 Results of Kirana Evaluation

Phase I and III

Figure 5 shows the mathematical scores from phases I and III for each of the participants. We observed an average improvement of 8.3% in mathematical test scores from phase I to phase III. However, four participants showed a decrease in performance. A score of 3 indicates knowledge of numbers from 1 to 10 and familiarity with money mainly notes, while a score of 6 indicates knowledge of numbers from 1 to 100, familiarity with notes and coins, and simple two-digit additions and single-digit subtractions. A score of 10 would ideally mean that the participant has knowledge of numbers from 1 to 100, is familiar with notes and coins, and is comfortable with mental arithmetic including single-digit multiplication. A paired-sample t-test of the math test scores for phase I and phase III of all 18 students shows no statistical significance ($p = 0.54$).

Phase II

A total of 218 items were purchased during the evaluation. During the first week, the participants spent an average of 9.3 minutes (SD = 4.5) per session buying items, while during the second and third weeks they spent 6 minutes (SD = 4.4) and 4.5 minutes (SD = 2.2), respectively. Consequently, participants spent an average of 2 minutes and 34s per item (SD = 1.39) during the first week, 1 minute and 43s (SD = 1.94) during the second week, and 1 minute and 13s (SD = 0.57) during the third week. This reduction in transaction time per item indicates that participants became comfortable with the transactions and interactions with the application. We note here the limitation in measuring task time; we cannot differentiate the time taken for decision making and that of the actual interaction. The participants spent an average of 71 rupees during the first week, 80 rupees during the second week, and 63 rupees during the third week.

Phase IV

Actual store visits in phase IV provided an ecologically valid way of translating the learning from virtual to real-world scenarios. Of the 18 participants, 12 went to a local store with an educator. Of the 6 participants that did not take part in phase IV, 2 had moved to another school and 4 were assessed as needing more practice with the application. It should be noted that this was the first time the 12 participants had come to an actual store by themselves. The educator observed the participants interact with the shopkeeper and buy items from a given list, similar to the task in phase II.
**Decision-making:** Participants were able to connect the tasks during the visit to the tasks in *Kirana* and purchase items from the given shopping list. The moderator observed a high degree of cooperation among the participants, and they helped each other out wherever possible. The most popular items during the visit were cold beverages due to the warm Indian summer.

**Social interaction:** Participants who were feeling shy would observe the others interact with the shopkeeper before approaching the store. One participant made the valid observation that everyone should take a bill from the shopkeeper so that the purchasing amounts can be checked by an adult later. Surprisingly, three usually nonverbal participants were very responsive and able to interact with the shopkeeper.

**Mathematical ability:** As expected, the arithmetic aspect of calculating costs and balance was highly dependent on one’s mathematical ability. Participants with a higher mathematical score (greater than 6) were comfortable with the concept of money and transactions involving arithmetic. Two participants gave all their money to the shopkeeper, one due to nervousness and the other had problems reading the decimal place in the price because of the small font size on the packet. All the participants were able to locate an item’s M.R.P. and understand that money is exchanged when purchasing the item, and they expressed a positive attitude towards purchasing. Several participants understood the relation between items and costs such that they only selected items they knew they would be able to purchase with the amount of money they had.

Overall, we observed improvements in mathematical ability, based on the teacher’s assessment, and translatable learning from the application to an actual local store. We believe *Kirana* did the necessary groundwork in preparing individuals with developmental disabilities for purchasing items of need from local stores, thereby teaching them a valuable life skill that promotes self-efficacy.

### 4 BALLOONS APPLICATION

*Balloons* [33] focuses on promoting the social activity of shared attention, or joint attention, between two users. It is important to motivate individuals with developmental disabilities towards such social interaction and collaboration as countermeasures against social isolation. Joint attention is defined as a shared experience over a common goal or object and is linked to language acquisition and social interaction in the later stages of the neurological development of a child [6]. For individuals with autism, encouraging joint attention is a crucial step towards social inclusion. Our goal for *Balloons* was to allow different explorations of the challenges in learnability of joint attention using gesture for selection pointing with dwell time.

#### 4.1 Design of Balloons

Similarly to the design of *Kirana*, we followed a user-centered design approach to design the application, involving 23 stakeholders: nineteen school specialists, 3 parents, and 2 high-functioning children with autism [33]. First, we conducted semi-structured interviews with the specialists, where they shared their own educational intervention guidelines as compiled over several years of hands-on work. Based on the common themes that emerged from the interviews, we conducted two focus group discussions. The end result is a list of design requirements for developing an application’s interaction, interface, and goals, which are as follows:

- Motivate and encourage social interaction even for those who have limited vocabulary, for example, via joint attention.
- Provide adequate rewards to encourage children to interact, regardless of their performance, for example, applause from the system.
• Provide a well-defined start and end of a session to allow the children to understand when a task is over.
• Encourage socially acceptable gestures, as gestures such as kicking or jumping, which might be fun for gameplay, can be dangerous and socially isolating outside of the session.
• Provide multimodal feedback, both visual and auditory, as children have varying levels of visual and auditory preferences, sensitivities, and capabilities.
• Include colorful graphics with smooth visual movements and illustrations to captivate attention for longer time spans.
• Empower the child by involving them in the decision making.

Designed to address these requirements, Balloons provides a simple way to promote joint attention via gesture-based interaction for individuals with developmental disabilities. It consists of three colored balloons (Figure 6, left) that can be selected by pointing and holding the hand steady on a balloon (dwell time) for 3s. An on-screen hand cursor shows the position of the hand on the screen in real time. Two users can interact with the application at the same time. If one user selects a balloon, then a star is rewarded (Figure 6, center). If both users select the same balloon together, then they are rewarded with a growing rainbow and pleasant music (Figure 6, right). The application aims to encourage social interaction between the participants. The use of a screen to mediate interaction between participants reduces the awkwardness of face-to-face communication or making direct eye contact, which is especially challenging for individuals with autism. The application was built on the same framework (Section 3.4) as Kirana.

An ideal Balloons session would require two participants to stand in front of the system facing the screen. As a first step, both participants would identify their respective on-screen cursor and then one would prompt the other to select a specific balloon together. The prompt can be verbal, by saying the color, or visual, by moving one’s cursor towards a specific balloon (both are instances of social interaction). The other participant would then also move his or her cursor towards that balloon. In this way, if both participants select the same balloon within the 3s dwell time, they are rewarded a rainbow (joint attention).

4.2 Balloons Participants
Fourteen individuals with a wide range of developmental disabilities participated in our evaluation (3 individuals with Down syndrome, 1 individual with high functioning autism, 1 individual with ADHD, 1 individual with cerebral palsy and an unspecified form of intellectual disability, and 8 individuals with other unspecified forms of intellectual disabilities). There were 10 females and 4 males aged between 18 and 41 years (M = 26, SD = 7). The participants were recruited from two different classes in the school:

Group one (n = 9), with an average IQ of 42 (SD = 11) and SQ of 41 (SD = 10). They (a) did not understand the concept of left or right, (b) communicated with a vocabulary of 50–150 words,
Fig. 7. (Left) Picture of the setup and (right) moderator guiding a participant by supporting the elbow.

(c) could not follow instructions, (d) had an awareness of self and body, and (e) did not participate in the Kirana study.

Group two (n = 5), with an average IQ of 63 (SD = 20) and SQ of 57 (SD = 15). They (a) did not understand the concept of left or right, (b) communicated with a vocabulary of 150–500 words, (c) could follow instructions, (d) had an awareness of self and body, and (e) did not participate in the Kirana study.

4.3 Balloons Evaluation Procedure

Balloons was set up in a classroom at Nai Disha School using a projector, a laptop, and a Kinect (similar to the one in Figure 7) for 10 days during February 2017. Each participant had two sessions with the application. The sessions were moderated by a male educator who is working on technology interventions at Tamana and who also moderated the Kirana study. However, the participants of this study were unfamiliar with him, and, therefore, group one’s special educator (female) was also present during all the sessions. This created a comfortable and familiar social environment for the participants, a majority of whom were also female. The participants in group one attended the sessions in pairs, i.e., in each session two participants would be present, where one participant interacted and the other looked on and vice versa. For group two, because of its small size, sessions were conducted with the entire group present at the same time. For both groups, each session was about 15 minutes long, with three breaks to reduce fatigue. The moderator frequently enquired whether the participant was comfortable and wanted to continue. The moderator initiated communication with each participant to build a rapport with them and asked them to name the colors of the balloons and whether they found the interaction interesting and enjoyable. In this way, the moderator made sure that the participants were willingly present.

First session: The aim of the first session was to confirm that each participant was comfortable using gestures for interaction and working with the moderator. Each participant was first asked to count the number of balloons they saw on the screen. Then they were asked to raise their right hand, during which the moderator stood behind the participant to provide support in case they experienced any fatigue. If they did, then the moderator would gently support the elbow of the participant and enquire if they were comfortable (similar to Figure 7, right). This was also repeated for the left hand. The moderator then stood in the front of the screen, in view of the Kinect, and demonstrated the pointing gesture using his right and then left hand. The participant was asked to repeat the same by standing in front of the screen and identifying their hand on the screen (right-hand cursor as shown in Figure 6), while the moderator supported them. The primary role
of the participant in the first round of sessions was to keep the hand steady for the 1.5s dwell time and select a total of five balloons.

Second session: In the second session, the participants were asked to independently select seven balloons. The moderator stood behind to support the participant by gently holding his or her elbow if he or she mentioned any fatigue.

Since the participants selected balloons with the moderator standing behind them, they were identified as one entity by the Kinect and a star was rewarded. The application was used as an introduction to gesture-based interaction, without focusing on its two-person interaction. However, we note that this usage of Balloons still encourages the social skill of joint attention, that is, a shared experience over a common object. Moreover, this allowed us to prepare the participants for the HOPE study, which required single-person interaction.

Data Gathered: We collected the total session time for each participant for each session, the number of balloons that were selected with assistance, and the number of balloons they selected by themselves. The moderator also noted observations after each session for each participant, namely, the type of assistance required and if there were any other challenges.

4.4 Results of Balloons Evaluation

As expected, group one required more time and assistance in completing the tasks during the two sessions with Balloons. To compare the task times between sessions, we normalized the data gathered with respect to the number of balloons selected per session, which were five in session one and seven in session two. We also define a completion score, which takes into account how many balloons were selected by the participant independently and how many required assistance.

Types of assistance

For each session, there were varying levels of assistance, which can be categorized as verbal or physical, with two types of verbal prompts, one for the content and one of the gesture. Physical assistance was provided by the moderator by standing behind the participant and supporting their elbow. Initially, this included supporting the elbow to guide the participant towards the appropriate content on the screen and to reduce interaction fatigue. For the subsequent sessions, the moderator supported the participants’ elbow to reduce interaction fatigue but provided no guidance. The type of assistance provided in each session is mentioned in Table 2.

Completion Score

We define the completion score (as a percentage) for each participant for each session to understand their performance at the task. This was calculated by

\[ \text{Completion Score} = \frac{(\text{number of balloons selected by self}) + 0.5 \times (\text{number of balloons assisted})}{\text{total balloons per session}} \times 100 \]

By giving less weightage for balloons that were selected with assistance, the completion score provides a useful metric to compare how well the participants are able carry out the gesture and

<table>
<thead>
<tr>
<th>Session</th>
<th>Verbal Assistance</th>
<th>Physical Assistance</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prompts for</td>
<td>To guide hand</td>
</tr>
<tr>
<td></td>
<td>Content</td>
<td>movement</td>
</tr>
<tr>
<td>Balloon I</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Balloon II</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

Table 2. Type of Assistance in Sessions
complete the task. Figure 8 shows the completion score for each group and session. As expected, Group 1 has a lower completion score than Group 2; in the first session for Group 1, all participants required assistance with the gesture for selection. While for Group 2, in the second session, all participants were able to complete the selection task without any verbal or physical assistance.

**Performance Time**
We calculated the performance time for each participant for each session. We define the performance time (in seconds) as the time taken for each individual task, i.e., time taken to select a balloon (individually or assisted). This was calculated as

\[
\text{Performance time} = \frac{\text{total time for a session}}{\text{total balloons per session}},
\]

where the total balloons selected for that session is equal to five for session one and seven for session two. With respect to performance time, Group 1 participants required more time for each session as compared to Group 2 (Figure 9). While both groups saw reduction in performance, the reduction was more drastic was Group 2.

**Observations**

*Group 1:* In the first session, seven participants found it difficult to follow the instruction to raise their right hand. They were all able to identify the colors blue, red, and purple. After the moderator re-demonstrated the instructions (raising right, left, and, finally, both hands), six participants were able to repeat them, while three could not. All participants were assisted by the moderator by gently supporting and nudging their elbows and holding their hand steady. It was also noted that participants took time to communicate with the moderator and build a rapport. Participants also
initially found it difficult to identify their hand on the screen. By the end of the first session, however, they were able to relate to the onscreen cursor.

In the second session, participants tried the task themselves, with the moderator occasionally supporting their elbow to reduce interaction fatigue. One participant faced particular difficulty in identifying their hand on the screen in the second session and then lost interest in the task. This was exacerbated by a system misbehavior where the application did not detect their hand correctly. Between sessions, participants seemed to open up to the moderator, as they would meet him in the school corridor and excitedly enquire about their next session.

Group 2: The sessions were conducted with all participants of group two present at the same time. The first session required minimal assistance and was completed in less than the 15 minutes assigned. However, one participant with autism did not show interest unless assisted. In the second session, participants were able to select seven balloons in under 2 minutes, where moderator assistance was only required by one participant who wanted to play with his/her friend (one of the observers). They selected balloons together and were awarded the rainbow.

When considered together, quantitative and qualitative data provide a holistic perspective of the potential of gesture-based interaction. Although the participants of Group 1 required more time and assistance, they still completed the tasks and were engaged in the learning process. We believe this feeling of accomplishment is more valuable and rewarding than focusing solely on the actual performance time.

5 HOPE APPLICATION

HOPE is an interactive learning application that promotes motor coordination and cognitive and social skills through a series of spatial reasoning tasks. It addresses two main challenges towards sustainable educational interventions in the Indian context: technology diversity and inclusive learning.

Technology diversity: Indian users have a multitude of old and new technology within the classroom and home environments. An application that supports multiple interaction mechanisms (e.g., mouse, touch screens/tablets, or gestures using a Microsoft Kinect sensor) can be used on a multitude of devices. This enables consistent and sustained usage of the application. For instance, a school with a Microsoft Kinect sensor can provide gesture-based interventions to their students within the classroom environment and parents can follow-up with at home sessions on a tablet.

Inclusive learning: Given the context of a special school, it is naïve to think one size fits all when it comes to educational interventions. For instance, within the autism spectrum, it is well known that needs between individuals can be quite diverse. Our previous gesture-based interventions therefore focused on a smaller subset of users, for example, Balloons for medium- to low-functioning children with autism [33] and Kirana for individuals with developmental disabilities who were previously unable to shop independently [34]. To be inclusive and cater to a large spectrum, HOPE provides multiple tasks with increasing levels of difficulty. Participants, parents, and educators are able to decide the level they want to start with.

Thus, HOPE provides a sustainable, consistent, and flexible educational intervention with multiple interaction mechanisms. Our research goal for this study was to explore the challenges in learning complex gestures for selection and object manipulation for individuals with developmental disabilities.

5.1 Design of HOPE

The application combines memory, spatial, motor, and cognitive skills based on two broad learning paradigms: matching shapes, colors, and objects, and the concept of numbers and alphabet.
Our focus in this article is on the matching tasks and using gestures for interaction. The matching tasks consist of three categories: colors, shapes, and objects. Each category consists of up to 20 screens with increasing levels of difficulty. For instance, for matching colors, the first screen consists of three items of the same shape: two green triangles and one red (Figure 10, left). From screen 6 onwards, the difficulty increases by having two items of the same shape and one of a different shape, as seen in Figure 10, center. From screen 11 onwards, another level of difficulty is introduced as all the items are of different shapes. Figure 10 right shows a black star and heart and a yellow circle, for which a participant has to match the star and heart as they are of the same color.

Similarly, for matching shapes, screen one starts with three items of the same color (Figure 11, left), while screen 6 introduces another color (Figure 11, center). From screen 11 onwards, the number of items increases from three to four (Figure 11, right). For the object-matching tasks, the first screen starts with familiar objects and shapes (Figure 12, left), while new shapes and objects are introduced from screens 6 and 11 onwards (Figure 12, center and right). Each correct response is rewarded with an auditory applause and flying balloons for visual stimuli.

The matching tasks require a sequence of gestures (grab → drag → drop) to complete. An ideal HOPE session will include a participant interacting with the onscreen content by solving the tasks presented using grab, drag, and drop gestures. For example, the first level of the shape-matching screen shows two blue squares and one blue triangle (Figure 11, left), and the following steps are required to correctly perform the matching task using gestures as follows:

1. Grab: A participant has to select one of the blue squares by grabbing, that is, start with an open palm (shown in Figure 13, left), move his or her hand on top of one of the squares, and then close the palm (shown in Figure 13, right). The blue square is then attached to the onscreen hand cursor.
2. Drag: With the palm closed, the participant has to move his or her hand to the other blue square. The onscreen hand cursor follows the movement path, dragging the selected square with it.
In this way a successful grab, drag, and drop gesture sequence is performed. The application starts with an instruction screen, as shown in Figure 14.

### 5.2 HOPE Participants

The participants from the Balloons study (Section 4.2) took part in the HOPE study.

### 5.3 HOPE Evaluation Procedure

There were a maximum of three sessions conducted with each participant using HOPE over the span of two weeks from February to March 2017. Similarly to Balloons, sessions with Group 1 were conducted in pairs and with Group 2 in a group. The task was matching colors, an extension to the task in Balloons. This task familiarity provided a base to build on, since the gesture (grab, drag, and drop) was now more complex. Verbal prompts were given throughout the sessions to help the participants with the gesture sequence and with the content. Physical assistance was provided to reduce interaction fatigue. The procedure for each of the sessions is now presented.

**First Session:** The first session included a practice session to provide a detailed explanation and demonstration of the grab, drag, and drop gesture. Unlike Balloons, here assistance was provided by the special educator who stood behind the participant. The special educator guided and supported the participant by gently nudging his or her elbow, similarly to the process followed by the moderator in the first session of Balloons (Figure 7, right). The participant and the special educator stood facing the screen and Kinect (similar to Figure 7, left). The moderator, however, stood next to the Kinect, with their back to the screen, facing the participant and out of the active area. This allowed the participant to see the moderator’s palm and follow the palm-open and palm-close instructions. The participant was asked to focus on the moderator’s hand movement and verbal prompts, while being guided by the special educator standing behind the participant. After practicing palm-open and palm-close several times, the complete grab, drag, and drop gesture was taught. The complete practice run and the accompanying verbal prompts are explained in Table 3.

After the going through the gesture sequence twice, the first color task screen was started. The session consisted of 10 screens for matching colors, as shown in Figure 10, for up to five different
colors: green, red, blue, yellow, and brown. The moderator and the special educator kept repeating the same gesture practice, that is, the participant could look at the moderator’s hand movement while the special educator gently guided the student’s hand to complete the onscreen task. Once the participant became comfortable with the gesture, he or she would start looking at the screen instead of the moderator, and at this point the moderator stopped repeating the gesture sequence. Most participants learnt the gestures by the fifth task screen and would then be able to interact with the content on the screen independently, without the special educator’s assistance. To reduce fatigue, instead of breaks, the participants were asked to exercise their hands by moving both hands up and down to relax the muscles.

Second and third sessions: In the second and third sessions, the participants were already familiar with the grab, drag, and drop gesture. Therefore, there was no elaborate explanation or demonstration of the gesture, and the moderator provided assistance by standing behind the participant. The moderator supported the participant’s elbow or hand, in case of interaction fatigue, while the special educator observed from a distance. The second and third sessions included the same screens as the first with an additional 11th screen, which was the most difficult and was added to keep participants, who were able to complete all 10 screens, engaged (Figure 10, right). These sessions were not run with Group 2 participants since they were able to comfortably and correctly complete the task in their first session.

Data Gathered: Similarly to Balloons, for both groups, we collected the total time per session per participant, the number of screens that were assisted, and number of screens completed by the participant independently.

5.4 Results of HOPE Evaluation

In the evaluation, the level and type of assistance played a significant role. After the practice gesture sequence, Group 2 participants completed the 10 task screens without any assistance. Meanwhile, Group 1 participants completed all the screens with some assistance in sessions one and two. However, because of the minimum level of assistance provided in session three, Group 1 participants completed between 5 and 11 screens with the longest performance time. Therefore, learning from session to session varied widely among the participants.

Table 3. Grab, Drag, and Drop Gesture

<table>
<thead>
<tr>
<th>Gesture</th>
<th>Moderator: action</th>
<th>Moderator: Verbal prompts</th>
<th>Special Educator: action</th>
</tr>
</thead>
</table>
| Ready   | Raises his right hand and then opens his palm | *Raise your hand*  
*Open your hand* | Gently nudges the participant’s elbow, moves hand over an object to select it on the screen |
| Grab    | Closes the palm  | *Close your hand*          | Supports the participant to select the object on the screen |
| Drag    | Moves closed palm sideways | *Move your hand*          | Moves the participant’s hand to the correct placeholder for the selected object |
| Drop    | Opens the palm   | *Open your hand*           | Supports the participant to correctly match the object |
Table 4. Levels of Assistance in HOPE

<table>
<thead>
<tr>
<th>Session</th>
<th>Assistance Verbal</th>
<th>Assistance Physical</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prompts for</td>
<td>To guide hand</td>
</tr>
<tr>
<td></td>
<td>Content</td>
<td>movement</td>
</tr>
<tr>
<td></td>
<td>Prompts for</td>
<td>To reduce</td>
</tr>
<tr>
<td></td>
<td>Gestures</td>
<td>fatigue</td>
</tr>
<tr>
<td>HOPE color I</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>HOPE color II</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>HOPE color III</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

Fig. 15. Completion score for HOPE.

Types of Assistance
Similarly to Balloons, for each session there were varying levels of assistance as described in Table 4. This assistance was mainly provided to Group 1 participants.

Completion Score
We define the completion score (as a percentage) for each participant for each session to understand his or her performance at the task. This was calculated by

\[
\text{Completion Score} = \left(\frac{\text{number of screens completed by self}}{100} + 0.5 \times \frac{\text{number of screens assisted}}{100}\right) \times \frac{\text{total screens attempted in that session}}{100}
\]

Figure 15 shows the completion score for each group and session. Since Group 2 participants were able to complete all screens in their first session, no further sessions were conducted with them. For Group 1 participants, the completion scores vary between 50 and 100%. For instance, a 50% completion score in session one means that the participant was able to complete all the screens but required physical and/or verbal assistance for each. In session three, Group 1 participants completed between 5 and 11 screens.

Performance Time
We calculated the performance time (in seconds) as the time taken to complete one screen (individually or assisted):

\[
\text{Performance time} = \frac{\text{total time for a session}}{\text{total screens completed in that session}},
\]

where the total number of screens completed in each session varied for each participant.

Figure 16 shows the performance time for the three sessions of Group 1 and one session of Group 2. Overall, Group 2 participants were quicker to learn the new gesture and complete the color-matching tasks as compared to Group 1. Moreover, in session three, Group 1 participants required more time to complete the gesture and task, as there was no verbal or guided physical assistance provided.
Fig. 16. Performance time for HOPE, with boxplots showing range, median, and quartiles.

Observations

Group 1: The moderator observed that by the fifth or sixth screen, the participants were able to familiarize themselves with the gesture sequence and could complete it with only verbal prompts, thus without the need to look at the moderator standing in front of them. By the end of the first session, participants were able to focus on the content of the screen. In the subsequent sessions, the gesture sequence was performed with verbal prompting (explained in Table 2), while the moderator stood behind the participant to support his or her hand to reduce interaction fatigue. Since the task content was similar to Balloons, the participants did not face any issues with executing it correctly. This made the prime focus of the sessions on learning the gesture sequence, as the content was considered relatable. However, we note here that most participants were new to the colors brown and yellow.

The grab, drag, and drop gesture in HOPE increased the complexity of the tasks, especially in comparison to Balloons. In Balloons, since the focus was on pointing, a half-opened palm did not cause any interaction issues. However, a known technical issue with the Kinect is that it recognizes an open palm as closed if it is below a certain height or facing away from the Kinect. Therefore, very specific hand movements were required to execute the gesture sequence: initiation with an open palm, selection by closing the palm, dragging by moving a closed palm across the screen, and dropping by opening the palm. Group 2 participants were able to understand this gesture complexity and complete the task in their first session. The technical limitation of the Kinect did not affect them. Group 1 participants faced challenges in performing the gesture sequence and maintaining a steady hand throughout their first session, which we believe was exacerbated by the technical limitation.

In their second session, Group 1 participants had a better understanding of the gesture and improved their performance time. However, in session three, not all Group 1 participants were able to complete the entire set of 11 screens and without verbal and physical prompts for the gesture or task required more time to complete each task screen. Therefore, Group 1 participants would benefit from more sessions with the color-matching tasks, while Group 2 participants can graduate to the next learning challenge, say, of matching shapes. This support for learning diversity in HOPE establishes the potential of designing an application to cater to a large group of individuals, that is, inclusion-within, one of main research goals of this work.

6 DESIGN GUIDELINES AND DISCUSSION

We presented three gesture-based applications and their evaluations with individuals with developmental disabilities in India, from investigating the use of pointing with dwell time for selection with two applications (Balloons and Kirana) to exploring the challenges in gestures for object
selection via grab and object manipulation via drag and drop in the HOPE application. Although complex, the gesture sequence of grab, drag, and drop for interaction was found to be learnable by individuals with developmental disabilities in the context of a simple problem-solving task (match objects by color). On the other hand, Kirana combined a complex set of tasks, including mathematical calculations, decision making, and social interaction, with simple pointing gestures. Our findings indicate that there needs to be a balance between task and gesture complexity, when employing gesture-based interaction.

In the following, we consolidate our findings as guidelines that highlight opportunities and help overcome challenges in designing gesture-based applications for individuals with developmental disabilities. We categorize the guidelines as those (a) specific to application design, (b) specific to research study design (i.e., ethical considerations and conflicts that might arise), and (c) specific to the Indian context. Furthermore, our guidelines adhere to the principle of *inclusion-within*, that is, designing applications that cater to diverse learners and nurture collaboration among users. Moreover, individuals with developmental disabilities form a wide spectrum of users, who, even when grouped under the same diagnosis, have diverse abilities and preferences. The guidelines aim to respect these differences and preferences when it comes to designing interactive applications, for instance, by providing multimodal feedback, having tasks that steadily increase in difficulty, and finding the right balance between gesture and task complexity.

### 6.1 Guidelines for Application Design

**A.I Provide a Clear Start and End of Gameplay**

It is extremely important to provide a clear start and end of gameplay to avoid ambiguity and confusion during interaction. For example, in Balloons and Kirana, when a participant enters the active area of the Kinect, i.e., the area most suitable for interaction, the screen turns white from black to reveal the virtual objects. Similarly, when the participant moves out of the active area, the screen turns black to indicate the end of interaction, that is, the application no longer responds to gestures. This can also assist in reducing excessive attachment with the application and discourage repetitive gestures and behaviors, a common concern when designing for individuals with developmental disabilities, especially autism [25]. Kirana also caters to various end-of-play scenarios such as not enough money left to buy an item or running out of money or items to buy. In each of these scenarios, the reason for the end is stated, followed by a visual and aural “well done” feedback to conclude the session. Therefore, each session with the application has a clear start and end, which provides structure to the interaction.

**A.II Provide Feedback with Multiple Modalities**

Individuals with developmental disabilities find it easier to select items that have both visual, e.g., a change in size, and auditory, e.g., name of the item said aloud, feedback. Feedback with multiple modalities helps overcome visual or auditory impairments, if any, and provides multiple stimuli for attention [2, 3]. In Kirana, each visual object, on being selected, also had auditory feedback. In HOPE, the moderator provided verbal prompts for gestures and content to assist the participants. Going forward, it would be better to incorporate the verbal prompts in the application with the option to control the volume. Overall, providing multimodal feedback allows for diverse stimuli that makes an application suitable for a larger group of participants.

**A.III Offer Rewards and Positive Reinforcements**

One way to encourage and motivate social or physical interaction is through continuous rewards and positive reinforcement. In Balloons, every selection is rewarded with a star (individual selection) or a growing rainbow with playful music (jointly selected). In HOPE, every correct answer is rewarded with flying balloons and an auditory applause. Physical rewards such as a high five,
clapping, and other harmless activities that are of interest and liking of the user can be used as rewards. A reward can also be a time-bound session with another favorite game to further maintain participant’s engagement level. Furthermore, it is important to add reinforcements like “shabash” (well done in Hindi), “well done,” or applause, regardless of the correctness of the task, to encourage and motivate participants to continue playing. Therefore, rewards and positive reinforcement should be designed for continuous encouragement and motivation.

A.IV Evolving Task Difficulty Levels with Easy Transitions
Applications should provide varying levels of difficulty with enough repetitions to allow the educator or caregiver to customize learning goals for each individual and cater to a large group of participants [2]. Moreover, it should be easy to repeat a level or specific task and to transition to the next level. For instance, as shown in Figure 10, in HOPE the matching colors task start with three items of the same shape: two green triangles and one red. From the 6th screen onwards, the difficulty increases by having two items of the same shape and one of a different shape. Then, from the 11th screen, another level of difficulty is introduced by having all items of different shapes. In this way, the level of difficult increases every 5 screens. Moreover, users can start from any screen, that is, there is no need to unlock the difficult levels. This enables the users to select content that is interesting and suitable, without having to go through disinteresting or disengaging content first. Therefore, applications that introduce increasing levels of difficulty in their tasks should allow for easy transitioning between levels and enough repetition of each level.

A.V Provide Serial and Structured Content
Designing content that is structured and follows a series of related steps provides a way to break down complex tasks into manageable steps. For instance, Kirana simulated the real-world scenario of buying groceries from a local store. This was achieved by designing interaction that is socially and culturally appropriate and by breaking down complex tasks that require social, mathematical, and decision-making skills. Moreover, the purchasing animations were slow and sequential. In HOPE, matching tasks are grouped by level of difficulty, with enough repetition of tasks for each level. Furthermore, designing serial and structured content also supports designing a clear start and end of gameplay and evolving task difficulty levels with easy transitions. We recommend complex tasks to be broken down into a sequence of smaller achievable steps. If the animation and other visual media content also follow the same step-by-step order, then individuals with developmental disabilities can follow the progress of the tasks with ease.

A.VI Simulate Real-World Scenarios
Gesture-based applications offer the potential of simulating real-world scenarios in a safe and controlled environment. This can be achieved in two ways: by incorporating socially and culturally appropriate interactions and by designing the content to match a real-world activity. Moreover, using free-form body gestures provides a mechanism for more inclusive social interaction and teamwork, especially when compared to touch screens that have limited form factor and surface area to support multiple users. This inherent nature of embodied interaction [38] makes for an immersive and engaging experience, which can be utilized to encourage real-world social interactions. In Kirana and HOPE, even though used by a single active user, the interaction was visible to the moderator and others who could then prompt and encourage the user, thus involving them in the activity. Overall, an application can be designed to simulate a real-world scenario through their content or interaction or both.

A.VII Balance Gesture and Content Complexity
Interactive applications incorporate two learning paradigms in one task—the content and the method of interaction. When designing an application with varying levels of difficulty, it is
important to make incremental changes to either the content or interaction method. For instance, in HOPE, the interaction gesture was the same for different tasks, such as matching color, shapes, or objects. In this way, the learning of the gesture sequence from matching colors is potentially translated to other more complex matching tasks. Moreover, the concept of selecting a color was translatable from Balloons to HOPE, when the gesture for the selection increased in complexity, from pointing with dwell time to grab, drag, and drop. Therefore, we believe that it is important to incrementally add to the task complexity, through either content or gesture, to allow for incremental learning.

6.2 Guidelines for Research Study Design

R.I Design for Diverse Learners for Inclusion-Within
A school environment caters to a diverse group of learners, separated either on the basis of chronological age or cognitive abilities, to provide learner appropriate goals. Likewise, gesture-based applications for individuals with developmental disabilities have usually focused on a narrow group of learners. However, Nai Disha School has identified a growing need to create a sense of unity between diverse groups of learners. With HOPE, we redefined inclusivity and designed an application that could cater to students within the whole school and nurture collaboration instead of competition. This provides an additional learning opportunity: how to accept the strengths and challenges of other students, i.e., inclusion within. Our guidelines for application design are also attuned to this goal.

R.II Promote Self-Efficacy
Applications that promote self-efficacy should employ gestures and interactions that are socially and culturally relevant and acceptable for the specific life skill. This is particularly relevant in the developing world, where social awareness about individuals with disabilities varies greatly. Moreover, the socio-cultural norms of a society dictate the degree of inclusivity and integration of individuals with developmental disabilities [4, 15]. Individuals with developmental disabilities, such as autism, display a tendency towards repetitive behavior, and teaching socially unacceptable gestures might increase their isolation, even if the gesture is fun during gameplay [2]. In this respect, Kirana is an application that is culturally relevant and socially acceptable to the community. We believe it did the necessary groundwork in preparing individuals with developmental disabilities for purchasing items of need from local stores, promoting self-efficacy.

R.III Address Socio-Ethical Challenges
Conducting research with individuals with developmental disabilities poses several ethical challenges. First, selecting a control trial is a challenge because of varying skill levels by age, diagnosis, and mood swings. Moreover, Tamana provides equal opportunities to all, therefore no one should be denied participation. Second, answering the research hypothesis might require a certain level of moderator restraint to collect comparable data. One example of this is the time taken per task. Moderator assistance or taking several breaks would not allow for comparable task times. However, it is more important to provide assistance to the participant when requested or required. In HOPE, our initial research aim was to look at the learning curve of the grab, drag, and drop gesture sequence, for which it was important to provide limited assistance. Yet, the moderator quickly realized the need for several levels of assistance and provided it. Third, participants have a tendency to get attached to a specific application, and they show an interest to continue after the study sessions are over, which Nai Disha School gladly facilitates. Therefore, any resolution of conflict between the research procedure and the participants’ interest should always be in favor of the participants.
R.IV Provide Assistance for Gestures
Using gestures for interaction can induce fatigue and require uncomfortable motor movements. Therefore, for individuals with varying motor and physical strengths, it is important to provide assistance to reduce fatigue. During the HOPE sessions, the moderator provided physical assistance to avoid gesture interaction fatigue. In the initial sessions, when the participants were learning the gestures, the educator supported their elbows and gently guided hand movements while standing behind them. Once a participant was comfortable with the gesture sequences, the educator continued to provide elbow support to reduce interaction fatigue. This can also be achieved by using a stand, which does not affect the motion recognition technology. It is thus important to provide support and assistance for interaction methods that require motor or physical exertion.

6.3 Guidelines for the Indian Context
I.I Design for Technology Acceptance
We followed a collaborative and stakeholder-inclusive approach for our work from Kirana to Balloons and HOPE. Since the applications were designed and developed with educators, therapists, and parents, who provided valuable insights and experience, there was no resistance towards technology acceptance. The applications were integrated within the classroom environment with ease. Additionally, we identified ways to address the challenges in introducing new technological interventions for individuals with developmental disabilities. First, economic barriers for technologies that are too costly can be overcome by designing applications that can be integrated within schools and can be personalized and customized for use by a larger group of individuals. Second, resource constraints can be overcome through collaborations among schools, universities, and industry partners. Third, the digital divide can be reduced by spreading awareness of the benefits of technology within schools and to parents. In India, gesture-based technology is rare and novel, and thus its acceptance remains largely unexplored. Our findings suggest that following a user-centered, collaborative, and stakeholder-inclusive design approach can reduce the challenges in technology acceptance.

I.II Address Technology Fears
During the course of our work at Nai Disha School, several parents enquired about the harmful effects of technology-mediated educational interventions from being addictive to possible sources of radiation. As researchers and practitioners who advocate technology, one needs to be very clear on what it means to use technology. For instance, we explain to parents that their children use technology at the school for 15–30 minutes a day and that the usage of technology is always monitored and moderated. Moreover, most of our interventions employ gesture interaction, which does not require holding of any devices. In fact, the Microsoft Kinect motion sensor works best if the participant is between 2 and 3m away from the device. Individuals with developmental disabilities, specifically with autism, get attached to games or devices, which can socially isolate them even further. Therefore, it is important to discuss technology apprehensions that parents and care givers might have, including technology addiction. This becomes especially relevant when there is potential for misinformation and misguidance around socially repressed topics, such as autism and developmental disabilities, which is common in the Indian context.

I.III Give Control to the Educators
One important way to increase technology acceptance is to provide control to the special educator. As discussed already, due to the wide range of abilities of the individuals, the educator or moderator should be able to customize the application, and its levels and screens, based on individual capabilities and interests of each user. While we believe providing control to the educator is important in the Indian context, it is universally applicable, as shown in previous research.
Moreover, in India, there is a particular need to personalize learning goals to cater to a large number of individuals and to increase technology acceptance and cost-effectiveness. For instance, in Kirana, the educator guided the social interaction and the task by giving a shopping list. This enabled the educator to customize the learning, even on the fly. Moreover, the number, type, and price of items, and the total amount of money in the wallet, could be changed easily between sessions. In HOPE, the educators had access to all the screens before working with the students, enabling them to understand the sequence and complexity of the task. This could help them identify the most suitable starting level for each student. Furthermore, when designing an inclusive and customizable application, a detailed curriculum of the tasks and learning goals can be incorporated as a guide. This guide can be made accessible from each screen with a small tab, something already available in the TOBY Playpad application [39].

7 CONCLUSION

This article presented three studies with gesture-based applications for individuals with developmental disabilities in India. The applications included Kirana, which teaches life skills to individuals with developmental disabilities by breaking down complex tasks; Balloons, which promotes joint attention skills; and HOPE, an application for improving motor coordination, joint attention, and cognition. Based on the exploratory user studies with these applications, we presented 14 design guidelines in favor of gesture-based systems for enabling educators and caregivers across the globe to impart social, motor, and life skills to individuals with developmental disabilities. Furthermore, our guidelines focus on applications that are inclusive and cater to a large group of individuals. This nurtures collaboration and helps to overcome challenges towards technology acceptance and adoption in India.

For individuals with developmental disabilities, technology provides a safe, controlled, and predictable learning environment, as shown by research all across the world. Gesture-based interaction provides an additional leverage by allowing for culturally appropriate social interactions and collaborations, reducing social isolation while also providing means for self-expression. Involving different stakeholders, for instance, parents, therapists, and special educators, in the design process reduces the barriers towards technology acceptance and adoption. Furthermore, applications that can be customized for different individuals, and diverse learning goals, cater to a large user group, enabling efficient use of resources. Going forward, we believe that our work and these guidelines pave the way for designing applications that can be used across schools in different cities in India and globally.

ACKNOWLEDGMENTS

Our thanks to the school for their valuable time and guidance, and a special thanks to all the participants. We also thank Sanna Grönlund for developing the graphics for the Balloons and Kirana applications.

REFERENCES


Received April 2017; revised November 2017; accepted November 2017
Paper V


Overcoming Socio-Technical Challenges for Cross-Cultural Collaborative Applications

Sumita Sharma¹, Pekka Kallioniemi¹, Tomi Heimonen², Jaakko Hakulinen¹, Markku Turunen³, and Tuuli Keskinen¹

¹Faculty of Communication Sciences, University of Tampere, Tampere, Finland
²University of Wisconsin-Stevens Point, Stevens Point, Wisconsin, USA
³Author Keywords
Cross-culture collaboration; collaborative online applications; interaction design.

ACM CCS
CCS → Applied computing → Education → Collaborative learning

INTRODUCTION
Cross-cultural online collaborations are increasingly common in higher education and the workplace. Therefore, it is desirable that schoolchildren also develop computer skills and conversational abilities in a foreign language [44]. However, inclusive collaboration, where children from different socio-economic and cultural backgrounds interact is not well studied. Moreover, children from underprivileged regions, defined as low socio-economic households with limited access to technology, are further excluded. Our research focuses on facilitating access to cross-cultural collaborative online experiences for schoolchildren from underprivileged regions. We started our work with collaborations between the researchers’ native countries – India and Finland.

Both Indian and Finnish students learn English as a foreign language in schools. In India, English language is seen as an important skill for employability and improving socio-economic status [10, 38]. The language barrier also deepens the technology barrier, as almost the entire web content in India is in English [20]. Furthermore, parents from low-income households aspire for both computer literacy and English for their children [38]. In Finland, English is considered a key to globalization in business, science, and it is the lingua franca for online and TV media, as well as rock and pop music [48].

In this paper, we present two user studies with underprivileged Indian students from Deepalaya. Deepalaya is a non-government organization, with schools or learning centers in the industrial areas of Delhi. Deepalaya is a non-government organization, with schools or learning centers in the industrial areas of Delhi, providing education to underprivileged students. The school has a computer classroom where students mostly learn how to use MS Paint, Word, or Excel. Using a laptop and an internet connection, we set up a collaboration with a Finnish researcher using the CityCompass application. CityCompass is a collaborative language learning application with 360-degree panoramic city views. The application provides hints and clues to support students with varying levels of English proficiency. The aim is to unite students with common educational goals in online environments, thus allowing collaborative learning between the developing and developed worlds, for example. We started our study with student-researcher pairs, due to the ethical concerns of pairing up young students without first understanding the challenges involved in the collaboration.

In our first study, we found that Indian students were hesitant to interact, as compared to Finnish students. Communication is a critical requirement to support collaboration and for conversing in a foreign language [13]. Earlier studies have shown that social and cultural norms dictate communication, which affects the learnability and usability of digital content.
The Indian students also faced technical challenges, due to the little to no access to computers.

Previous studies with Indian adults suggest that the Bollywood Method, where tasks are rooted in dramatized stories, can overcome cultural and social inhibitions to communication [4, 5]. Inspired by the Bollywood Method, in our second study we added a short scenario to the application task, in order to answer the question: Can the Bollywood Method reduce socio-technical challenges to collaboration for underprivileged Indian students in an online learning environment?

The novelty of our work is two-folded. First, we adapted the Bollywood Method to a different context – from Indian adults being encouraged to think aloud in usability tests (original context) to Indian children communicating in a foreign language. Second, we connected children from underprivileged backgrounds to individuals outside of their local and national context for one-to-one collaborative online tasks. The rest of the paper starts with related work, including a description of the Bollywood Method. Next, we describe the methodology of our user research and the CityCompass application. Finally, we present our results and discuss the significance of scaffolding cross-cultural collaborations with dramatized scenarios in enabling global collaboration.

**BACKGROUND**

We first situate our work within research on technology for children in India and online cross-cultural collaboration. We then describe the Bollywood Method and compare it with other scenario/narrative-based design methods.

**Technology for Children in India**

Low functional literacy, lack of access to resources, and a lack of material in local languages has resulted in a large digital divide, especially in developing regions [41]. However, most of the current research is focused on rural adults interacting with non-textual or semi-textual UIs [11, 29, 30, 31]. Research with rural Indian children includes the ‘Hole in the Wall’ studies by Mitra et al. [32, 34, 35], Kam et al.’s work on designing digital games for English language learning on mobile phones [21, 22, 23, 24], and speech-based English language learning applications [27, 33]. Our research extends prior work by enabling global collaborations.

**Using Mobile phones:** Studies show that mobile educational gaming is especially suited for technology deprived rural India; it provides an engaging, immersive and rewarding experience for new skill acquisitions [1, 20, 21, 22, 23, 24, 26]. Kam et al. [22, 23] successfully developed culturally relevant mobile games for out-of-school rural Indian children. Their focus was primarily on mobile phones catering to children who are unable to attend schools due to financial and practical constraints; the child is also a working member of the family, financially supporting the other members [20, 22]. The games are usually single player or require limited peer collaboration.

**Using Computers:** Children from underprivileged regions in Hyderabad showed considerable improvements in verbal English skills by using a speech to text software for three months [33]. This is part of Mitra’s ongoing efforts towards Self-Organized Learning Environments, SOLE [8], for children from disadvantaged backgrounds. It also includes the “school in the cloud” where rural Indian children video chat with “Grannies”1 or teachers from all over the world for various educational goals. The granny cloud is similar to our work, in that, it enables cross-cultural collaboration. However, in our work, students use a computer, which promotes hands-on computer skills and the collaborations are mainly one-on-one.

**Online Cross-Cultural Collaboration**

Modern web technologies offer a platform for knowledge sharing and distribution in a global context across cultures [12]. Yang et al. [51] suggest that these technologies should be utilized to enable and improve communication and collaboration. Moreover, when conversing in a foreign language, children also greatly benefit from social interaction and collaboration [13, 47, 49], especially during online gameplay [42].

Previous studies on online learning environments report that culture subtly guides communication and interaction, which has an impact on learning and collaboration [46]. Based on previous work on online collaborative learning [43, 46] and work with Indian adults [4, 5], two main aspects of culture are particularly relevant for communication – power distance and face-saving. Power distance refers to the expectation and acceptance of unequal social status or power distribution within a community. In societies with a large power distance, such as in India, there is unequal social status or a strong hierarchy among people [15, 16]. With respect to learning, communication is usually initiated by an authority figure who cannot be contradicted out of respect within and outside the classroom environment [46].

Face-saving is a universal social phenomenon deeply rooted in culture and the concept of identity [2, 43]. It refers to the tendency of avoiding conflicts within social communication. With respect to learning, face-saving, where the credibility and reputation of both child and teacher is not threatened or questioned, is further strengthened [7, 46]. Both face-saving and power distance are understood to affect communication, inter and intra culturally. To overcome the influence of these inherent cultural and social norms on communication we looked towards adapting the Bollywood Method.

**The Bollywood Method**

While conducting usability studies with Indian participants, Chavan [4] found that users would not share negative feedback out of politeness, and were hesitant in admitting the

1 http://thegrannyccloud.org/
problems they faced. The Bollywood method was devised by Chavan to overcome the socio-cultural challenges in communication within the Indian culture—namely *power distance* and *face-saving* [4, 5]. Bollywood, or the Hollywood of India, creates movies that usually have a larger-than-life fantasy. In the Bollywood Method, users are presented with a dramatized scenario that requires them to take on the role of a character with a specific goal. For example, for testing an airlines ticket booking application, users were asked to imagine that their niece is unknowingly getting married to an underground hit man, who is already married. The users must book tickets to Bangalore with the incriminating evidence in their sole possession to stop the wedding [4]!

The Bollywood Method is different from other scenario-driven usability evaluation methods where the user is asked to imagine themselves in a realistic scenario related to the product usage. In the Bollywood method, the scenario is dramatically heightened—one that would *typically* not occur. The scenario is also grounded specifically within a culturally relevant narrative and projects a very personal experience. Studies with Indian adults, with varying levels of literacy and technology experience, show that the Bollywood Method [4, 5, 6, 29] overcomes cultural inhibitions allowing users to communicate more openly and provide critical feedback. It has been successfully used with both rural and urban Indian communities. However, these studies are focused on uncovering usability issues and not on collaboration in online environments for children.

In our view, schoolchildren are influenced by the same social constraints—*power-distance* and *face-saving*. Furthermore, children with limited computer experience face both social and technical challenges towards collaboration. We hypothesize that a Bollywood Method inspired scenario can potentially overcome the social-technical challenges towards online collaboration. This would allow learning through cross-cultural collaborations among children from different parts of the world, with varying computer and language skills.

**Other Scenario-based Approaches**

There are several methods to engage children within a technology design or research process, with varying levels of involvement. From end users of an available application, as in our case, to testers and informants during the development processes, to design partners [9]. In our work, children participated as testers, they were observed and direct feedback on their experiences with CityCompass was elicited in order to improve the application in the future. These methods are grounded within the larger framework of participatory design (PD) and scenario-based methods [3, 24, 36, 37], and are suitable for motivating children’s involvement in learning activities [14, 40]. Our adaptation of the Bollywood Method can be considered a special case within the larger domain of scenario or narrative-driven design process and digital storytelling.

Within scenario/narrative-driven design, our method slightly resembles the concept of forum theatre [36, 37], where a short play is used to improve “communication through the use of embodied (i.e., acted-out) experience and through contextualized narratives” [37]. Furthermore, in their work on traditional rural games, Kam et al. [22] observed that South Indian children performed a short play or *skit* before starting their game. This performance of a scenario before gameplay closely resembles our adaptation of the Bollywood Method, that is, providing a background scenario to a game-task only before starting that task, and not during.

Digital storytelling can “increase students’ understanding of curricular content and improve their technical, collaboration, and communication skills as they engage in long-term storytelling projects” [40]. Moreover, it provides “diverse, personalized, compelling, and realistic” ways for children to practice and learn using technology [14]. This may not be the case in general for scenario/story based learning technologies for children.

**METHODOLOGY**

To test our hypothesis of the viability of Bollywood method on cross-cultural communication, we conducted two user studies with a between-subjects design: Study 1 served as the control group and in study 2 (experimental group) students were introduced a dramatic scenario (independent variable) at the beginning of the task.

In both the studies, students from Deepalaya in Delhi collaborated with a native Finnish researcher, who connected from Tampere[2]. Using CityCompass, student-researcher pairs communicated in English to complete a wayfinding task situated in the city of Tampere. Sessions at Deepalaya were moderated by a native Indian researcher. Both the Finnish and Indian researchers have spent considerable time in India and Finland, respectively. Given their experiences with the two cultures, they were well suited to observe subtle social-cultural nuances in communication that could affect the collaboration.

**Data Collection**

For both studies, we collected automated system logs from CityCompass, self-reported demographic information and user experiences, moderator observations, and interview data for each participant. Automated logs recorded system interaction data, which included the time spent on each panorama and in total on the route, number of hotspots activated, number of dead-ends and number of pans during the task. Before using CityCompass the participants self-reported their age, grade, months of computer experience (including usage at home or classes within or outside of the school) and years of studying English as a second language.

---

2 https://youtu.be/Jbn2bAIxPcU (demo video)
After the session, quantitative user experiences were gathered with a questionnaire and a semi-structured interview was conducted. The questionnaire consisted of 15 statements answered on a five-step Likert scale ranging from Totally disagree to Totally agree and having Neither agree or disagree in between (refer to Figure 8). The questionnaire did not use the popular Smileyometer [39] because Indian children have a tendency to choose smiles over frowns because of their aesthetic value [21]. Seven of the statements were from the SUXES method [45] and inquired the speed, pleasantness, clarity, ease of learning, naturalness, usefulness and willingness of future use of the application. Statements concentrating on the effortlessness of using the application correctly and overall liking of the application were also inquired. The bases for the UX questionnaire and interview have been developed over several years of user studies with novel interaction methods [e.g., 25]. The moderator conducted the interview in English and Hindi and it included questions about what the participants found to be the best and worst features in the application, and what they would like to improve. Regarding the statement 'I believe I learnt something today’ in the UX questionnaire, the participants were also asked what they thought they had learnt by using the application.

Participants
In the first study, 25 participants (M=4, F=21) were recruited, aged between 9 and 18 years (M=11, SD=2) and with an average of 5 years studying English as a second language in school (SD=3 years). For three participants this was the first time they interacted with a computer, while for two other participants it was their 4th or 5th time. Overall, computer experience varied from none at all (n=3) to 24 months (M=7 months, SD=8). Computer experience here refers to roughly how long a participant is enrolled in Deepalaya’s computer course, which mainly consists of classes two hours per week where students learn how to draw (MS Paint), write (MS Word), and enter data in Excel. Of the 25 participants, 23 did not have a computer at home, and one participant’s neighbour had a computer which they could use. 14 participants had no computer gaming experience, while 11 occasionally played some computer games during their computer lessons in school (refer to Table 1).

The second study was conducted with 22 participants (M=9, F=13) aged 9–18 years (M=12, SD=2) who had been studying English for an average of 4 years (SD=3 years). These participants were not a part of study 1. Eleven participants were using a computer for the first time; overall computer experience varied from none at all (n=11) to 24 months, but two students reported 4 and 5 years (M=9, SD=18). 21 out of the 22 participants did not have a computer at home, and only 9 participants had previously played a computer game during their computer classes in school (refer to Table 1).

<table>
<thead>
<tr>
<th>Table 1: Participant demographics for study 1 and 2.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study 1</td>
</tr>
<tr>
<td>Study 2 (scenario)</td>
</tr>
<tr>
<td>Total number of participants (Male, Female)</td>
</tr>
<tr>
<td>Mean Age in years (SD)</td>
</tr>
<tr>
<td>Mean Years studying English (SD)</td>
</tr>
<tr>
<td>Mean Months studying Computers in School (SD)</td>
</tr>
<tr>
<td>Number of participants who are first time computer users</td>
</tr>
<tr>
<td>Number of participants with computer access at home</td>
</tr>
<tr>
<td>Number of participants with some previous gaming experience</td>
</tr>
</tbody>
</table>

CityCompass
CityCompass is a language learning application, where two users communicate using Voice over IP connection. It supports three interaction paradigms (a) traditional mouse and keyboard using a web interface, (b) bodily interaction (via the Microsoft Kinect sensor), and (c) a head mounted virtual reality system (via the Samsung Gear VR headset). For this work, we used the traditional mouse and keyboard based interaction. This requires a computer with an internet connection, which is available at Deepalaya computer rooms.

In CityCompass, a tourist and a guide work collaboratively to reach a preassigned destination in a city. The route in the application consists of a sequence of seven 360 degree panoramas of an actual city, in our case, Tampere, Finland. Both users navigate the same panoramas. However, each user has her own view of the panorama which she can pan freely using click and drag. Each panorama contains several informative landmarks, called hotspots. On mouse hover, a hotspot’s informative text and accompanying audio is played. Each panorama has up to four exits, which are visible to the tourist as green arrows. Of these exits, only one is correct. The guide sees the correct exit as a blue line. Selecting the correct exit takes both users to the next panorama. An incorrect exits take them to a dead-end panorama in which the tourist needs to describe her location and the guide needs to select the correct location out of four possibilities, to get back. Therefore, at each point, for dead-ends and exits, the tourist and guide have to collaborate to reach the destination, requiring verbal communication. An example hotspot (mouse on the trash can) and green arrow (possible way forward) for a tourist is shown in Figure 1. The guide’s view with the route marked by a blue line, is shown in Figure 2.

Procedure for Study 1 and Study 2
Participants for both studies were recruited with the help of their teachers and on a volunteer basis, mainly from grade
levels 7th and 8th. Similar to Kam et al. [20], we found it difficult to accurately define an educational baseline as participants’ English reading and speaking abilities varied widely within the same grade level. Furthermore, participant ages also varied within the same grade level, since several students started schooling late or had gap years. Therefore, the moderator would ask the participants their name and age in English in order to gauge whether they would be able to understand the guide. The moderator also asked them to read aloud the application instructions at the beginning of the session (Figure 3).

Participants used CityCompass during their school day and were not compensated for their time. None of the participants were familiar with CityCompass or the city of Tampere. The physical setup was at two separate locations: in a classroom at Deepalaya for the students (tourist), as shown in Figure 4, and the other for the researcher (guide) in Finland. Each setup consisted of a laptop with a wireless mouse, an internet connection, and headphones with a microphone for the verbal communication. We did not use video calling in any of the sessions because of the low internet bandwidth available in the school area. The first panorama was considered as practice and its data was excluded from our analysis, although this was not explicitly mentioned to the participants.

The evaluation started with the moderator providing a brief description of CityCompass to the participant and explaining the role of a tourist. Participants were also told that there is no time limit for completing the wayfinding task. The participants were asked to read the CityCompass instruction screen aloud and to practice panning the panorama and clicking on the image of the green arrow (Figure 1). The briefing was provided in English and Hindi (the languages used at the school). Hindi was used to make sure the students completely understood what they were participating in and they would feel comfortable to ask questions in Hindi, if not in English. Each participant completed the wayfinding task only once and participated only once.

During the sessions, the participants’ teachers were not present in order to reduce the large power distance that might affect the participant’s behavior. In India, teachers are addressed as Ma'am and are considered an authority figure that cannot be contradicted or spoken to without permission [46]. The Indian moderator was introduced to the participants as didi, a respectful way to address an elder sister in Hindi, which implies a lower level of authority than Ma'am. The Finnish researcher was introduced by his first name, as a peer. After finishing the task, by reaching the destination, the participants filled in the UX questionnaire and the interview was conducted.

A similar study with Finnish students has been conducted before. In this, Finnish students interacted with an Indian researcher using the CityCompass route of New Delhi (instead of Tampere). The study included 19 Finnish participants (M=9, F=10) from grade 3 (age M=12, SD=1) with an average computer experience of 6 years (SD=1) and three years of English language classes. 18 out of the 19
participants had a computer at home; 14 of them played games regularly and 4 played occasionally. There was a considerable difference in interaction behavior and task completion times between the Indian students in study 1 and Finnish students of this study. The differences can be attributed to (a) computer experience: Finnish participants’ previous computer and gaming experience carried through to the interaction with the application, that is, they approached it like a game with a sense of urgency (shorter task times), with frequent panning and hotspots activation. (b) Social communication: by virtue of the generally individualist Finnish culture, which causes a relatively flat power-hierarchy, students were comfortable in interacting with the Indian researcher, and not hesitant in using a foreign language [46].

Adding the Bollywood Method in Study 2
In the second study with the Indian students, a scenario was introduced to create a sense of urgency in the wayfinding task and to reduce socio-cultural barriers towards collaborations. The procedure remained the same for both studies, except for the addition of a scenario at the beginning of the gameplay, after participants read the instructions screen, and practiced panning, in the second study. Participants were asked to read out loud a text document, in easy to read colloquial Hindi as shown in Figure 5, and translated below it.

We must note here that the story needs to be culturally relevant to excite its audience. Based on interactions with the students in the study 1, we found that the students did not travel alone or in groups to school trips, especially not outside of the country. Moreover, low-income families did not tend to travel for leisure. Therefore, we believe that the concept of a school trip to a foreign city and getting lost alone has dramatic appeal for our user group, and this inspired our story.

![Figure 5: The Bollywood Story in Hindi.](image)

**Translation:** You are on a school trip in Tampere with your class and the teacher. Somewhere along the way you got lost from the rest of the group and you are alone in an unknown place. You have to find your way from the railway station to the Laakontori market quickly before your group leaves. Fortunately, one of the locals has agreed to help you to find your destination. Let’s go!

**RESULTS**
In the following, the results from the second user study are compared to those from the first user study.

**Task Completion and Interaction**
We observed a statistically significant reduction in task times (in minutes), per an independent samples t-test, between study 1 ($M=34, SD=14.5$) and study 2 ($M=24, SD=8$); $t_{38.942} = 2.992, p<0.01$. There were no statistically significant differences between the average number of dead-ends, average pans per minute or the participant demographics (age, computer and gaming experience and years of studying English) between the two studies. There was however a difference in the average number of hotspots activated per minute; $t_{1.129} = -7.334, p=0.001$, that is, participants in study 2 used the available hints within the CityCompass more frequently.

We further divided the participants in study 2 into two subgroups – the first subgroup consisted of the eleven students who had no computer experience and the second subgroup consisted of the remaining eleven students who had some computer experience. An independent samples t-test showed no statistically significant difference between these subgroups for gameplay interaction with the application (number of dead-ends, pans per minute, hotspots per minute and overall task times). Thus, in study 2 the eleven students using a computer for the first time performed similar to the eleven students who had some computer experience.

Given that the experience of English is similar for participants of both studies and that study 2 participants had overall less computer experience than those in study 1, the differences observed between the task times and average number of hotspots activated per minute appear to have resulted from applying the Bollywood Method, as intended.

**Self-Reported Learning**
In the interview, the moderator also asked the participants what they believe they learnt while using the application. It is interesting to note the self-reported learning data, as it represents the participants’ attitude towards the application and its personally understood benefits. We coded the responses into four subcategories, which we grouped into two main categories; English language learning and computer learning. We defined English language learning as responses that mentioned the word English or having a conversation in English, asking for directions, talking to the guide, and understanding guide’s instructions. For example, a participant in study 2 said that she “learnt how to talk in English to a person from another country”. We defined computer learning as responses that mentioned how to use the system, system interactions and gameplay mechanisms, such as, learning to find their way in a city, clicking on the green arrow, and turning left or right. For example, a participant in study 1 exclaimed she “learnt how to play a game on the computer and to identify routes and paths to reach a destination.” Several participants mentioned both English and computer related learnings in their responses, as one participant mentioned, that she learnt to “speak in English and play a computer game.”
In study 1, 67% of the responses indicated English language learning and 33% computer learning. In study 2 English learning was 90% of the responses and computer learning was only 10%. A Chi-Square test (with 2x2 contingency matrix) shows a statistically significant difference between the studies and the response category (English and/or Computer): \( \chi^2 (1, N=55) = 4.731, p<0.05 \). We believe that the Bollywood Method assisted in moving the explicit learning of how to interact with the system to tacit or incidental learning. Most of the participants in study 2 believed they learnt how to understand and converse in English, the prime motive of the application, and were less aware of learning how to use the system. This is particularly interesting given that there were eleven participants who were using a computer for the first time in study 2 compared with only three in study 1.

**Communication Analysis**

From the perspective of language learning, the two user studies are too short term for any strong conclusions. However, the researchers observed that as participants conversed with the guide while using CityCompass, they gradually became more comfortable. Although communication analysis is not within the scope of this particular work and the audio recordings contain a lot of background noise from the school, we decided to analyze a part of the communication. Therefore, we transcribed the last five minutes of participants’ continuous (comfortable) conversation with the guide that ends with the participant reaching the desired destination in CityCompass. We did this for five randomly selected participants from each study. We categorized their utterances/words as confirmation (saying or agreeing in one word, for example, yes or OK), replies (answering a question asked by the guide), asking (asking a question to the guide), new-vocab (using a word that is not a part of the vocabulary used by the guide, moderator or hotspot information) and negations (saying no). The average number of utterances by category per participant are shown in Figure 7.

As expected, we observed increased communication (almost double the average number of confirmations, replies and questions) from the participants when using the Bollywood Method in study 2. An independent samples t-test also showed a statistically significant increase in the number of responses for: replies (study 1 M=7.8, SD=1.6; study 2 M=14.4, SD=2.8); t(8) = 4.45, p<0.05, and new-vocab (study 1 M=0; study 2 M=1.6, SD=1.5); t(8) = 2.359, p<0.05. In the five audio samples of study 1, none of the participants used a new vocabulary word and one participant said ‘no’ once. While in the five audio samples of study 2, all expect one participant used a new vocabulary word (twice on average) and similarly three participants said sentences that had a negative reply (again, twice on average). For example, one participant corrected the guide while another asked the guide to ‘please repeat’. This shows that the use of Bollywood Method helps overcome face-saving during negotiations and comfortably question an authority figure. An example conversation from study 1 and study 2 is shown in Figure 6.

**User Experience**

Figure 8 shows the average user experience questionnaire ratings given by the participants in the studies. Most of the ratings show a positive experience of using the application and there are no significant differences between the two studies. Thus, user experience remained consistent even with the addition of the Bollywood Method. For the UX question - hints offered are annoying - many participants specifically mentioned that they did not find hints annoying per se but that the audio-hints sometimes overlapped with the guide's conversation. This made it difficult to listen to and understand the guide. Moreover, during the sessions we observed that participants repeated the audio hints when replying to the guide, even if they did not understand the words. This shows participants' interest in listening to the audio hints and explains the high positive rating of the UX question - hints are useful.

**Finnish Researcher’s Observations**

In both the studies, several participants were really hesitant and/or shy about the task, but after a while they started to proceed along the route with ease. In wayfinding tasks, the most basic concepts are that of left and right, and some participants confused left for right, or vice-versa. This could be due to the too abstract nature of the application, the lack of English skills, or the lacking guidance of the researcher.
For some participants, the concept of a dead end or going back to previous panorama was difficult to understand. Most of the participants utilized the hotspots and used them constantly as reference points while trying to find the correct route. Many of these problems are wayfinding related issues, and therefore out of the scope of this particular work. The addition of the story in study 2, increased the participants' activity extensively; they started performing the task much quicker after the introductory phase. This could be due to higher level of presence created by the immersive story.

The first challenge we observed was towards communication. Students displayed a need for authoritative confirmation and a reluctance to answer negatively to the Finnish researcher. This can be attributed to the universal phenomena of face-saving [2, 7], where “neither the teacher nor any student should ever be made to lose face” [46], and the large power distance that usually exists in Asian cultures. For collaboration, this puts the onus of communication on the teacher or authority figure. Chavan noticed similar socio-cultural barriers towards communication in her work on usability of online services with Indian adults. To overcome this she devised a method to create a sense of urgency and provide a role or character for the user [4, 5].

Adapting the Bollywood Method from Indian adults [4, 5] to children and from usability evaluations to cross-cultural online collaboration, in the second study, we added a scenario around the wayfinding task, creating a sense of urgency. The scenario was provided in plain text form, just before the students started using the application. This sense of urgency provided (a) (as expected) a socially acceptable ‘face’ for participants to critique and contradict authority figures, thereby reducing the power distance, and (b) surprisingly, created game-like elements that reduced the implicit focus on using a computer. Furthermore, in study 1, one third of the student responses to what do you think you learnt? was using a computer, while this dropped to 10% in study 2. This is even more surprising given that the number of first time computer users was higher in study 2 (eleven students) as compared with study 1 (three students). Therefore, it can be stated that students in study 1 were implicitly aware of ‘using a computer’ during the session, which probably shifted the focus from collaborating with the researcher.

Furthermore, in other work with CityCompass [18, 19], students with considerably more technology access and exposure, used the application like a game. They seem to have a sense of urgency to finish the task and reach the goal. This game-like approach towards collaboration seems like a direct result of technology access and more individualist societal norms, when comparing overall Finnish or Northern European culture with Indian/Asian.

The success of the Bollywood Method, as a scenario to provide a specific context and agency (and urgency) to an online collaborative task, can be understood by comparing it with other similar scenario-based approaches for promoting user participation. This includes Muller’s descriptions of forum theatres as a participatory design and scenario-based
method [36, 37] and the strikingly similar theater or skit performance of rural South Indian children before starting their traditional (non-tech) games [22]. Furthermore, within the domain of participatory design methods for designing technology with and for children, several similar scenario, narrative, and storytelling tools are used [14, 40]. Although, the exact tool used depends on the extent of children’s participation – the level of involvement and available time – that is desirable in the design process [9].

Our scenario was developed based on interactions with study 1 participants who found it strange and shocking that the panoramas in CityCompass are of an actual city and not just random images (like realistic paintings). Upon further enquiry from the teachers at Deepalaya, we learnt that students did not travel outside of Delhi for school trips, or with their families, unless it was to their village. We understand that the actual scenario might not seem overly dramatic, especially from a western perspective, where schoolchildren travel on group trips lasting several days. This is also common in (expensive) private Indian schooling. However, given the conditions of families and children at Deepalaya, where children attend free government schools and go to Deepalaya for remedial classes, the used scenario is in fact fairly dramatic. Unfortunately, in an urban slum, traveling for leisure is not an option.

We note here that the cultural and social context of collaboration with a researcher may or may not be comparable with collaboration with a peer. Thus, face-saving and authoritative social interaction can be in part due to the collaboration between a student and an older researcher who would take the social role of an elder sibling or teacher. However, we argue that our results should easily extend to peer-collaboration due to the universality of face-saving, also among peers [7, 43]. Moreover, in typical learning environments, the presence of an authoritative figure is almost always expected, and learners should be supported in both social situations – when collaborating with a peer or with an adult.

Overall, the scenario reduced the socio-technical barriers towards cross-cultural collaboration between Indian students and a Finnish researcher. We believe our findings extend to other cultures and regions. The social barrier is not only common to Indian culture, but is prevalent in most Asian cultures [15], and generally in societies with large power distance. Face-saving is understood to be universal [43, 46]. Moreover, the digital divide is not a binary construct with a visible line between the haves and have-nots, but rather a complex socio-economical stratification within and among both developed and developing countries [28, 50]. Therefore, it stands to reason that the Bollywood Method can be adapted to benefit online environments for cross-cultural collaboration for children across the world.

**CONCLUSION**

Our work focuses on reducing the socio-technical challenges towards cross-cultural collaboration in online environments.

In this paper, we presented two studies with underprivileged children in India, who have little to no access to computers and currently lack opportunities for creating a global perspective and presence. We observed that when Indian students initially interacted with a Finnish researcher to complete a way-finding task using CityCompass, they were hesitant to communicate. We incorporated the Bollywood Method to CityCompass, by adding a dramatic story, and conducted a second user study. Our results indicate that the Bollywood Method (a) reduced the socio-cultural barriers to communication, (b) improved digital interaction within CityCompass, and (c) changed the focus of self-reported learning outcomes from *how to use a computer* to *how to communicate in English*. The results suggest that adapting the Bollywood Method can potentially improve cross-cultural collaboration. We believe our results can be extended to other domains, besides wayfinding task, and cultures, allowing children from different cultures and backgrounds to experience the globalized world we all share.

**FURTHER RESEARCH**

We adapted the Bollywood Method to introduce a sense of urgency by presenting a dramatized scenario before beginning an online wayfinding task. Such scenarios need to match the application domain, be appropriate for the students’ culture, and be relatable for the students even if somewhat dramatic [5]. The most suitable way to provide information depends on the students’ cultural background and their literacy levels. We chose to provide the scenario as a written text document which was read out aloud by the students. It would be interesting to note the level of immersion and engagement with different scenario presentation methods, such as animation, video or even a simulated telephone conversation. Furthermore, a variant of the scenario could be used to reduce the sense of urgency in students by presenting a different scenario. This is especially desirable for students who are either only interested in completing the wayfinding task as quickly as possible or prefer individual play, instead of meaningful collaboration.

CityCompass itself could be adapted to incorporate encouragement or guidance to increase communication. For example, if there is no verbal activity or interaction with the application for a significant duration, or the correct exit has been visible for a long time, the application could provide suitable conversational cues to initiate a discussion. Furthermore, panoramic views of actual cities have the potential to promote the experience and learning of cross-cultural artifacts and environments, providing further content for communication during cross-cultural collaborations.

**ACKNOWLEDGMENTS**

We thank all the students who took part in the user studies and their teachers for supporting this work.

**SELECTION AND PARTICIPATION OF CHILDREN**

Before the work started, a consent form was signed by Deepalaya HR allowing the researchers to introduce technology-mediated interventions to their schools. Teachers
in the specific school were given a detailed presentation explaining the research, application, procedure, and a mock session with the Finnish researcher. During the study, the teachers asked students in their classroom to volunteer. Interested students agreed on a session time with the moderator. At the beginning of each session, students were briefed and asked to read aloud the instruction screen. The ones who did, were asked if they would like to interact with a Finnish researcher. The others were asked if they wanted to play, and then 'played' informally with the moderator (these data not included in the studies presented).

REFERENCES


Publications in the Dissertations in Interactive Technology series

Details of the dissertations are available at http://www.uta.fi/sis/tauchi/dissertations.html

1. **Timo Partala**: Affective Information in Human-Computer Interaction
2. **Mika Käki**: Enhancing Web Search Result Access with Automatic Categorization
3. **Anne Aula**: Studying User Strategies and Characteristics for Developing Web Search Interfaces
4. **Aulikki Hyrskykari**: Eyes in Attentive Interfaces: Experiences from Creating iDict, a Gaze-Aware Reading Aid
5. **Johanna Höysniemi**: Design and Evaluation of Physically Interactive Games
6. **Jaakko Hakulinen**: Software Tutoring in Speech User Interfaces
7. **Harri Siirtola**: Interactive Visualization of Multidimensional Data
8. **Erno Mäkinen**: Face Analysis Techniques for Human-Computer Interaction
9. **Oleg Špakov**: iComponent – Device-Independent Platform for Analyzing Eye Movement Data and Developing Eye-Based Applications
10. **Yulia Gizatdinova**: Automatic Detection of Face and Facial Features from Images of Neutral and Expressive Faces
11. **Päivi Majaranta**: Text Entry by Eye Gaze
12. **Ying Liu**: Chinese Text Entry with Mobile Phones
13. **Toni Vanhala**: Towards Computer-Assisted Regulation of Emotions
14. **Tomi Heimonen**: Design and Evaluation of User Interfaces for Mobile Web Search
15. **Mirja Ilves**: Human Responses to Machine-Generated Speech with Emotional Content
16. **Outi Tuisku**: Face Interface
18. **Joel S. Mtebe**: Acceptance and Use of eLearning Solutions in Higher Education in East Africa
19. **Jussi Rantala**: Spatial Touch in Presenting Information with Mobile Devices
20. **Katri Salminen**: Emotional Responses to Friction-based, Vibrotactile, and Thermal Stimuli
21. **Selina Sharmin**: Eye Movements in Reading of Dynamic On-screen Text in Various Presentation Formats and Contexts
22. **Tuuli Keskinen**: Evaluating the User Experience of Interactive Systems in Challenging Circumstances
23. **Adewunmi Obafemi Ogunbase**: Pedagogical Design and Pedagogical Usability of Web-Based Learning Environments: Comparative Cultural Implications from Africa and Europe
24. **Hannu Korhonen**: Evaluating Playability of Mobile Games with the Expert Review Method
25. **Jani Lylykangas**: Regulating Human Behavior with Vibrotactile Stimulation
26. **Ahmed Farooq**: Developing Technologies to Provide Haptic Feedback for Surface Based Interaction in Mobile Devices
27. **Ville Mäkelä**: Design, Deployment, and Evaluation of Gesture-Controlled Displays in Ubiquitous Environments
28. **Pekka Kallioniemi** Collaborative Wayfinding in Virtual Environments
29. Sumita Sharma Collaborative Educational Applications for Underserved Children: Experiences from India