Available online at www.sciencedirect.com

SciVerse ScienceDirect

Int. J. Human-Computer Studies 71 (2013) 803–820

International Journal of Human-Computer Studies

www.elsevier.com/locate/ijhcs

To TUI or not to TUI: Evaluating performance and preference in tangible vs. graphical user interfaces

Oren Zuckerman*, Ayelet Gal-Oz

Sammy Ofer School of Communications, The Interdisciplinary Center (IDC) Herzliya, P.O. Box 167, Herzliya 46150, Israel

Received 2 March 2012; received in revised form 21 March 2013; accepted 9 April 2013
Communicated by Anatole Lecuyer
Available online 25 April 2013

Abstract

Tangible user interfaces (TUIs) are often compared to graphical user interfaces (GUIs). However, the existing literature is unable to demonstrate clear advantages for either interface, as empirical studies yielded different findings, sometimes even contradicting ones. The current study set out to conduct an in-depth analysis of the strengths and weaknesses of both interfaces, based on a comparison between similar TUI and GUI versions of a modeling and simulation system called "FlowBlocks". Results showed most users preferred the TUI version over the GUI version. This is a surprising finding, considering both versions were equivalent in regard to most performance parameters, and the TUI version was even perceived as inferior to the GUI version in regard to usability. Interviews with users revealed this preference stemmed from high levels of stimulation and enjoyment, derived from three TUI properties: physical interaction, rich feedback, and high levels of realism. Potential underlying mechanisms for these findings and practical implications are discussed.

Keywords: TUI; GUI; Tangible; Physical; Graphical; Interaction design

1. Introduction

Since the early days of personal computers, the Windows, Icons, Menus, Pointing device (WIMP) interface, invented at Xerox PARC, has been dominant in many of the digital devices we use. WIMP interfaces, also referred to as Graphical User Interfaces (GUI), have remained the dominant interaction model in personal computers and in mobile devices, despite the fact that many other interfaces have been explored during the past two decades. To name a few: touch-based interface, gesture-based interface, voice-based interface, and tangible user interface (TUI), which is the focus of the current paper.

1.1. TUI

TUI is a type of user interface that leverages physical representation to connect between the physical and the digital worlds. TUI is a field of research within Human Computer Interaction (HCI), and has seen increasing interest among HCI researchers in the past two decades (Shaer and Hornecker, 2010). TUI was first explored by Fitzmaurice et al. (1995), who presented their seminal work on "Graspable User Interfaces", using their "Bricks" prototype to present three key ideas: (1) physical artifacts which act as handles for control, (2) the advantage of leveraging people's lifelong experience with the physical world, and (3) space-multiplexed vs. time-multiplexed devices. The term "Tangible User Interfaces" was coined by Ishii and Ullmer (1997), who defined TUI as a "new kind of HCI… coupling digital information to everyday physical objects and environments" (p. 235). They presented a series of prototypes and suggested a classification of TUI to three classes: Interactive Surfaces, Coupling of Bits and Atoms, and Ambient Media.

Following this early work, a range of prototypes have been developed in the TUI domain, including TUIs for learning, programming, problem solving and entertainment, for example: AlgoBlocks (Suzuki and Kato, 1995), Digital Manipulatives (Resnick et al., 1998), Electronic Duplo Blocks

---

*Corresponding author. Tel.: +972 9 9527609; fax: +972 9 9527650.
E-mail addresses: orenz@idc.ac.il (O. Zuckerman), goayelet@idc.ac.il (A. Gal-Oz).

1071-5819/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ijhcs.2013.04.003

(Wyeth and Purchase, 2002), SystemBlocks (Zuckerman and Resnick, 2003), Topobo (Raffle et al., 2004), BodaBlocks (Buechley and Eisenberg, 2007), Tangible Programming (Horn, 2008), Media Blocks (Ullmer and Ishii, 1999), Block-Jam (Newton-Dunn et al., 2003) and many more. TUI systems have been classified and studied by HCI researchers (e.g. Dourish, 2001; Hornecker and Buur, 2006; Rogers et al., 2002; Shaer and Jacob, 2009; Ullmer and Ishii, 2000), and some TUI prototypes have even been commercialized (e.g. Sifteo, see http://www.sifteo.com; Topobo, see http://www.topobo.com).

1.2. The "TUI advantage"—theoretical explanations

Several HCI researchers suggested that tangible user interfaces have added advantages over graphical user interfaces. The pioneers of the field (Fitzmaurice and Buxton, 1997; Fitzmaurice et al., 1995) emphasized the bimanualism and space-multiplexed advantage of TUI vs. the time-multiplexed nature of GUI, and the advantages of natural affordances in tangible objects. More recently, Marshall (2007) claimed that TUI has great potential to support learning due to its "hands-on" nature, which allows physical manipulation of objects. However, he highlighted the need for additional evidence to validate the utility of TUI.

Klemmer et al. (2006) drew from psychology, sociology, and philosophy in order to formulate five theoretical themes explaining the importance of physical elements in interaction design. The first theme, *thinking through doing*, describes how thought and action are inherently integrated and together can produce learning and reasoning. The second theme, *performance*, describes how actions are faster and more nuanced compared to symbolic cognition. The third theme, *visibility,* describes the role of artifacts in collaboration and cooperation. The fourth theme, *risk*, explores how the uncertainty and risk attributes of physical co-presence shape interpersonal and human–computer interactions. The fifth theme, *thickness of practice*, suggests that embodied interaction is a more prudent path. Based on these themes, Klemmer et al. encourage interaction designers to integrate the computational and physical worlds.

Further support for the assumed "TUI advantage" can be found beyond the HCI literature, in psychological and educational research, demonstrating how various forms of physical interaction can enhance memory, performance and learning. For example, Hecht et al. (2008) demonstrated superior performance once a haptic signal was added to visual and audio signals. Participants in their study were able to detect the tri-modal combination (visual–auditory–haptic) faster than any of the bi-modal combinations, which in turn were detected faster than any of the uni-modal signals.

The same research group (Hecht et al., 2005) also highlighted the advantage of the visual–auditory–haptic combination in establishing a greater sense of presence in virtual environments. The authors hypothesized that the underlying cognitive mechanism is related to faster mental processing of multimodal events. The tri-modal combination enables users to start their cognitive process sooner, thus, in a similar exposure time, they can pay attention to a wider range of details and subtle cues. The integration of informative cues from different sensory modalities results in a richer and more coherent experience, which in turn leads to a greater sense of presence.

Moreover, gesturing was found to improve memory (Stevanoni and Salmon, 2005) and learning (Broaders et al., 2007). Goldin-Meadow et al. (2009) suggested that gesturing facilitates learning by helping learners extract information from their own hand movements.

In sum, various theoretical explanations, both within and beyond HCI literature, imply that tangible interfaces should be superior to graphical interfaces in regard to performance and learning. Accordingly, several researchers attempted to empirically demonstrate this presumed "TUI advantage".

1.3. The "TUI advantage"—empirical findings

Numerous studies compared between TUI and GUI or more generally between physical and digital interactions. Table 1 summarizes the main studies, their design and key findings[1].

As can be seen in the table, previous studies comparing TUI to GUI differ from one another in several ways. In regard to target population, some focused on children (e.g. Cheng et al., 2011; Manches et al., 2009) while others on adults (e.g. Marshall et al., 2010; Patten and Ishii, 2000). Overall, the majority of participants in comparative studies were children. Previous studies also differ from one another in regard to the specific research method employed, ranging from naturalistic observations (e.g.: Horn et al., 2009; Marshall et al., 2009) to highly controlled experiments (e.g.: Tuddenham et al., 2010; Xie et al., 2008). A key challenge of any comparative research is balancing conditions. In trying to balance between GUI and TUI conditions, unique affordances of one type of interface might get constrained or even entirely eliminated. Consequently, the benefits of the interface might get eliminated as well. As a whole, the existent body of knowledge overcomes this challenge by employing different research methods in different studies: some researchers employed a highly controlled experimental design, attempting to balance between TUI and GUI conditions to allow an accurate comparison. Others preferred a less controlled design, so the unique affordances of TUI and GUI could be explored.

In regard to evidence for the existence of a "TUI advantage", prior work remains inconclusive. While some studies suggest that TUI has a performance advantage over GUI (Fitzmaurice and Buxton, 1997; Tuddenham et al., 2010; Xie et al., 2008), others suggest it does not (Horn et al., 2009). Similarly, while some studies suggest TUI enhances cognitive functioning (Patten and Ishii, 2000), others were unable to demonstrate any advantages in the context of learning (Fitzmaurice and Buxton, 1997; Marshall et al., 2010). In addition, while several studies show that TUI is more inviting and engaging than GUI (Horn et al., 2009), others did not find any significant differences in

---

[1] A similar summary table was presented by Cheng et al. (2011). The current table is updated and extended to include additional studies and details regarding method and participants.

Table 1
Summary of prior work comparing tangible and graphical user interfaces, organized chronologically.

Authors	Interfaces compared	Study design and participants	Dependent measures	Findings
Fitzmaurice and Buxton (1997)	TUI GUI	A within-subjects experimental design $n=12$, age not specified	Performance Ease of use Learning rate	TUI outperformed GUI in regard to performance and ease of use, but not in regard to learning rate.
Patten and Ishii (2000)	TUI GUI	A between-subjects experimental design $n=36$ adults	Recall Spatial encoding	TUI outperformed GUI in regard to recall and sophistication of spatial encoding techniques employed.
Jacob et al. (2002)	TUI GUI Physical	A within-subjects experimental design $n=13$ adults	Task completion time Preference	TUI outperformed physical and GUI in regard to both task completion time and preference, though only at weak statistical significance levels.
Terrenghi et al. (2007)	Physical Multitouch	A within-subjects experimental design $n=12$ adults	Task completion time Satisfaction	Physical outperformed multitouch in regard to task completion time and satisfaction.
Xie et al. (2008)	TUI GUI Physical	A between-subjects experimental design $n=132$ children	Task completion time Engagement Enjoyment	TUI and physical outperformed GUI in regard to task completion time and engagement, but not in regard to enjoyment.
Manches et al. (2009)	GUI Physical Pictorial No materials	A series of within-subjects studies $n=12$–80 children	Problem solving	Physical representation outperformed GUI, pictorial representation and no material in regard to the problem solving strategies employed.
Horn et al. (2009)	TUI GUI	A between-subjects naturalistic observation $n=156$ adults $n=104$ children	Invitation Apprehendability Collaboration Engagement Performance Child-focus Invitation	TUI outperformed GUI in regard to being more inviting and promoting collaboration and child-focus, but not in regard to performance, apprehendability or engagement.
Marshall et al. (2009)	Physical multitouch	A qualitative video analysis $n=30$ children	Control mechanisms	More subtle control mechanisms were employed in the physical condition.
Tuddenham et al. (2010)	TUI Multitouch Mouse and puck	A within-subjects experimental design $n=12$ adults	Task completion time Error rate Comfort Preference	TUI outperformed Multitouch and Mouse and puck in regard to task completion time, error rate and preference rates, but not in regard to ratings of comfort.
Marshall et al. (2010)	GUI Physical	A between-subjects experimental design $n=34$ adults	Learning	No significant differences in learning.
Cheng et al. (2011)	TUI GUI	A between-subjects experimental design $n=32$ children	Engagement	No significant differences in engagement.

engagement (Cheng et al., 2011) or enjoyment (Xie et al., 2008) between the two interfaces.

Due to these inconsistencies in prior findings, one cannot reach an evidence-based conclusion regarding the “TUI advantage”. Therefore, a more promising direction would be to examine the context in which each type of interface offers more benefits. For instance, it appears from prior findings that TUI is more inviting than GUI and can better facilitate collaboration among users (Horn et al., 2009). These findings suggest TUI would be preferable to GUI when the task at hand is not inviting enough on its own, or when several users must work together in order to complete it.

A similar conclusion was reached by Horn et al. (2012), who focused on tangible interaction and learning amongst children. They defined the distinctions between GUI and TUI as "subtle", noting that one type of interface might be better for certain situations or for certain children, but not definitively better than the other.

Even if adopting this approach of matching interface to context, we still need to explain the contradictions between prior findings of studies comparing TUI to GUI. We believe these contradictions are, at least in part, a result of two forms of methodological concerns. First, as mentioned above, a key challenge of any comparative research is balancing conditions, which might end up constraining or even eliminating the unique affordances of each interface. Different researchers control for different benefits in each study, and as a result, different findings are obtained in each study. For example, consider the study by Xie et al. (2008), which compared children's enjoyment while playing with GUI and TUI versions of jigsaw puzzles. The authors explained how they wished to conduct a "valid comparison", therefore "the TUI puzzle was implemented to include the same modalities of feedback (auditory and visual) and available operations as the GUI puzzle" (p. 193). No significant differences in enjoyment were found. One can only speculate whether different results might have been obtained if the TUI version provided haptic feedback in addition to the auditory and visual feedback. Similarly, Jacob et al. (2002) compared between TUI, GUI and physical versions of a platform for organizing information. They tried to give the GUI version nearly all the properties of the TUI version, resulting in "an interface that was not quite a conventional mouse-and-keyboard GUI" (p. 344).

Secondly, contradictions in the findings of prior studies might reflect the variance in research methods employed by different researchers. As mentioned earlier, some researchers compared TUI and GUI in a controlled lab setting (e.g.: Marshall et al., 2010; Terrenghi et al., 2007), while others compared them in the field (e.g.: Horn et al., 2009). Some researchers used mainly quantitative measures (e.g.: Marshall et al., 2010; Xie et al., 2008), while others used mainly qualitative measures (e.g.: Marshall et al., 2009). Some researchers compared similar versions of the same system (e.g.: Horn et al., 2009; Marshall et al., 2010), while others compared systems that differed from each other in several key aspects beyond the type of interface (e.g.: Tuddenham et al., 2010). In addition, most studies, especially those involving adult participants, were based on a relatively small sample, which makes it particularly hard to draw any definite conclusions.

A notable exception is the study by Horn et al. (2009), which consisted of observing museum visitors interacting with either a graphical or tangible programming interface. The researchers observed a large number of users and used a mixed method approach: a wide range of comparative measures, but also interviews with selected participants. While certainly a comprehensive study, it was not experimental, therefore it does not enable us to infer causality between interface type and performance. In contrast, we decided to conduct a controlled experiment which would allow us to infer causality. We compared between similar tangible and graphical versions of a system called "FlowBlocks"—a custom-built computational simulation tool.

1.4. FlowBlocks—TUI

FlowBlocks is a modeling and simulation environment (Zuckerman and Resnick, 2003, 2004, 2005), which is a simplification of the System Dynamics modeling language (Senge, 1990; Sterman, 2000).

FlowBlocks, originally developed by the first author at MIT Media Lab, are wooden blocks with embedded computation that connect to one another using magnetic connectors, and allow users to form models of "data flow structures" (see Figs. 1–3).

When a model is formed, users can simulate the flow using light "tokens" that are passed from block to block. Lights can be transferred in different rates and passed through blocks with different behaviors, resulting in a range of different simulations (Zuckerman and Resnick, 2004). For the current study, we defined fives basic types of models, each representing a different type of simulation in the FlowBlocks language: (1) Chain no outflow—representing a rate of flow that increases accumulation, for example rain (as flow) and level of water in the lake (as accumulation), (2) Chain with outflow—representing a rate of flow that increases accumulation, and a rate of outflow that decreases accumulation, for example rain (as flow), level of

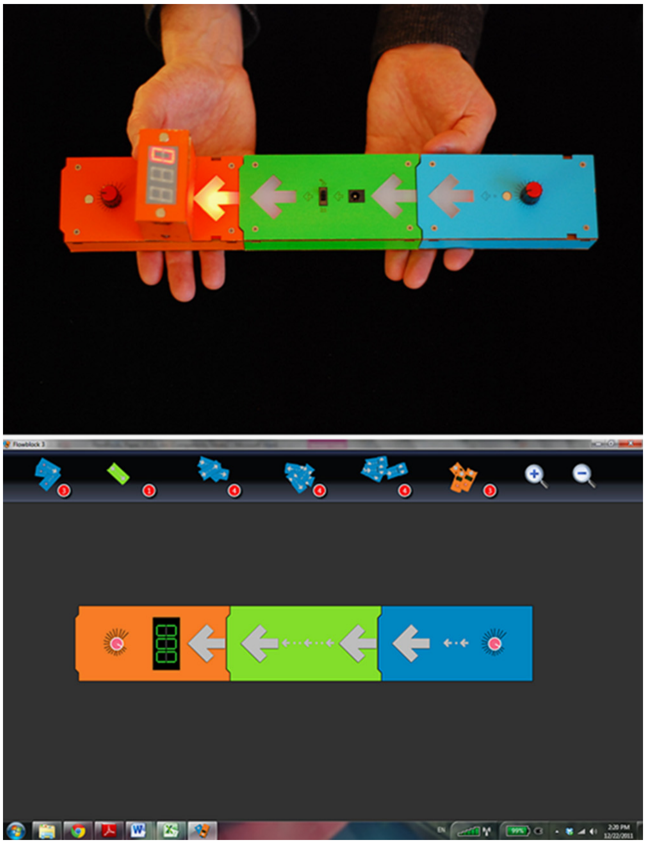

Fig. 1. FlowBlocks TUI version (top figure) and GUI version (bottom figure). The model is a simple "chain no outflow" 3-blocks model.

water in the lake (as accumulation), and rate of evaporation (as outflow), (3) Chain with distributer—representing a rate of flow that splits to two flows based on the distributer value, for example garbage in the landfill (as flow), 20% of it decomposes and 80% of it stays in the landfill as solid waste, (4) Chain with two accumulators—representing a rate of flow that increases accumulation, a rate of outflow that decreases the first accumulation, but leads into a second accumulation, for example rain (as flow) and level of water in the lake (as accumulation), rate of water directed from the lake to the fields (as outflow), and level of water watering the fields (as second accumulation), (5) Closed loop—representing flow in a closed system with no waste, can be with or without accumulation, for example the water cycle.

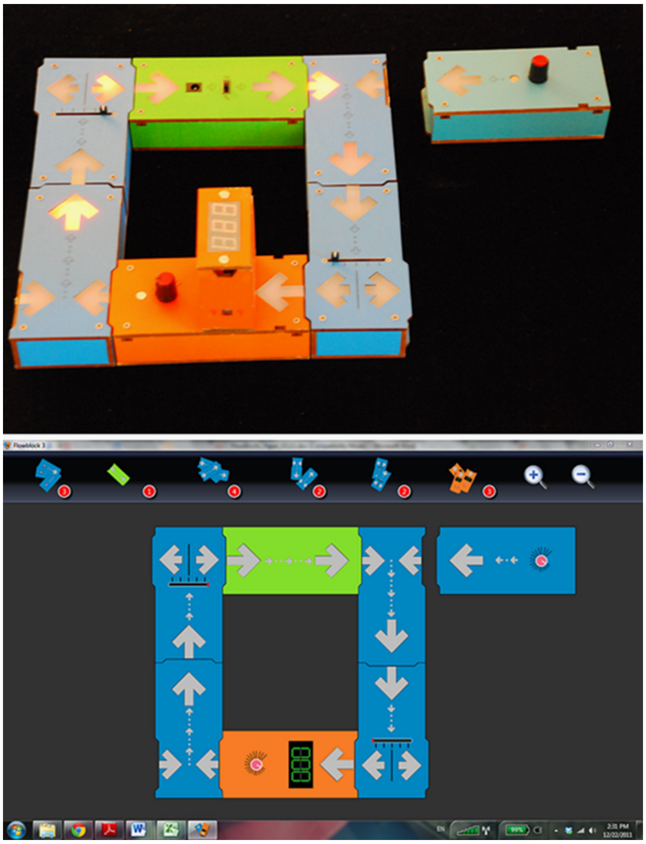

Fig. 2. FlowBlocks TUI version (top figure) and GUI version (bottom figure). The model is a cycle model, simulating flow and accumulation in a "closed loop" model.

Previous studies with FlowBlocks (Zuckerman et al., 2006) and its predecessor prototype "SystemBlocks" (Zuckerman and Resnick, 2005) showed that (1) SystemBlocks can help researchers map the various misconceptions children have about System Dynamics concepts, and (2) interacting with FlowBlocks affects children's thinking progression along the Structure–Behavior–Function framework (Hmelo et al., 2000; Hmelo-Silver et al., 2003). Misconceptions regarding system dynamics concepts are not limited to children; they are evident in adults as well. Booth Sweeney and Sterman (2000) found that even elite business school students demonstrated a poor level of system dynamics understanding. Furthermore, level of understanding was not related to prior education, age, national origin, or other demographic variables. Therefore, it appears that adults and children alike could benefit from using the FlowBlocks system.

In the current study, we compared between similar TUI and GUI versions of FlowBlocks, looking at the interaction aspects rather than at the learning experience. We aimed to strike a balance between conducting an accurate comparison and maintaining some of the unique affordances of TUI and GUI.

1.5. FlowBlocks—GUI

In order to compare between similar TUI and GUI versions of the same system, we developed a GUI-based version of FlowBlocks specifically for this study, making sure that the two interfaces are as similar as possible, with an identical set of blocks and behaviors (see Figs. 1 and 2).

TUI-based interaction and GUI-based interaction are different, so we had to design UI widgets to support the interaction with the GUI system. For example, turning a block in the TUI version is done in a natural way, using one or two hands to physically rotate the block. In the GUI version we created a "Turn Button" icon on each block, allowing users to turn a desired block by clicking on it (see Fig. 4, left).

Another example is the "Disconnect Block" action. In the TUI version, users use a natural action with two hands to disconnect one block from the other. They hold one block with one hand, place the other hand on the other block, and slowly

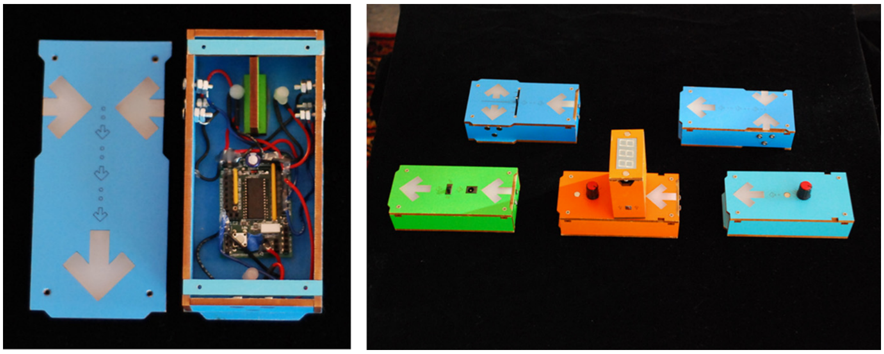

Fig. 3. FlowBlocks (TUI version), the electronics components inside a FlowBlock (Left), and the family of blocks (Right): left-bottom "power block", middle-bottom "Accumulator Block", right-bottom "Continuous Flow Block", left-top "Distributor Block", right-top "Turn Block".

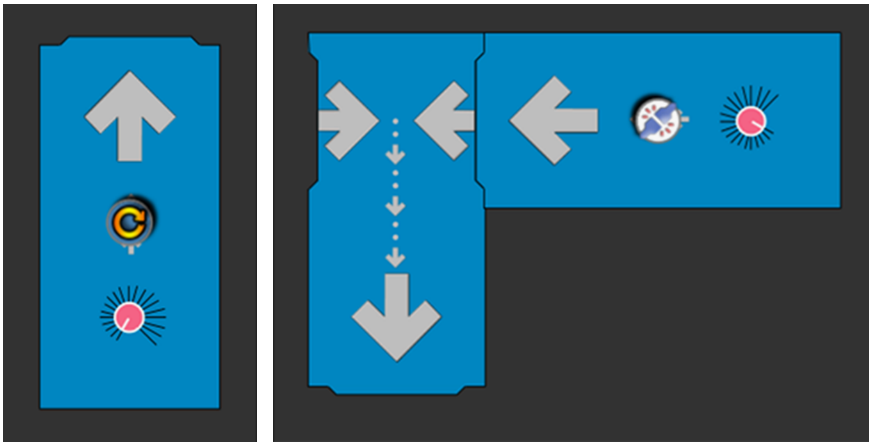

Fig. 4. FlowBlocks (GUI version), the special UI widgets and icons developed especially for the GUI version, to mimic TUI natural interaction operations. Left—the “Turn” icon to turn a block. Right—the “Disconnect” icon to disconnect one block from the other.

pull both blocks away from each other, until the magnetic connectors have been separated and the blocks disconnected. In the GUI version, we created a “Disconnect Block” icon, which appears in the middle of a block when a block is connected (see Fig. 4, right), and users can click it to separate one block from another.

We conducted a short pilot test to make sure that the newly created GUI version of FlowBlocks is easy to use with clear affordances. As expected, users who were unfamiliar with FlowBlocks found it easy to use.

While we attempted to balance between the TUI and GUI versions of FlowBlocks, several differences between the two versions still remained. The first difference between the TUI and GUI versions of FlowBlocks stems from the different kind of connectors used in each version. In the TUI version, blocks are connected to each other using custom-made magnetic connections that transfer power and data signals from block to block. Due to this type of custom-made-connectors, connection errors may occur when simulating models, but re-connecting blocks and/or restarting the system's power supply usually solves the problem. Such connection errors do not occur in the GUI version. Secondly, the blocks in the TUI version are more visible than the blocks in the GUI version, because TUI blocks are actual 3-D objects requiring a relatively large amount of space, whereas GUI blocks are 2-D graphical representations confined within the space of a standard computer monitor.

While these differences between the TUI and GUI version of FlowBlocks compromise the balance between the two experimental conditions, they represent natural differences between tangible and graphical interfaces: GUI is often more reliable than TUI, whereas TUI is often more visible than GUI. In other words, while we aimed to balance between the TUI and GUI versions of FlowBlocks, we did not “control out” all their respective affordances.

1.6. Overview of current study

The current study set out to extend prior research comparing tangible and graphical interfaces, by evaluating the advantages and disadvantages each interface holds in regard to users' performance and preference.

In order to be able to infer a causal relationship between interface type and users' performance and preference, we opted to use an experimental study design. We conducted the study in a laboratory setting, which enabled us to control for additional variables that might affect users' performance and preference. Specifically, we made sure the comparison between TUI and GUI was based on two similar versions of the same system, and that standardization was maintained for all participants.

Furthermore, in light of previous studies showing gender differences in response to TUI systems (e.g. Horn et al., 2009; Xie et al., 2008) and in regard to technology acceptance in general (e.g. Venkatesh and Morris, 2000), we included gender as a possible moderating variable in the experimental design.

In order to conduct as comprehensive analysis as possible, we filmed all experimental sessions to allow a careful examination of actual performance, as well as included several self-report measures to evaluate potential preference for one version of FlowBlocks over the other.

1.7. Research hypotheses

Based on the framework offered by Fitzmaurice and Buxton (1997), emphasizing the bimanualism and space-multiplexed advantage of TUI vs. the time-multiplexed nature of GUI, we expected users' performance would differ between the TUI and GUI versions of FlowBlocks. When working with the TUI version, users can utilize both hands, whereas only one hand is active when working with the GUI version (moving the cursor). Furthermore, the space-multiplexed nature of the TUI version enables users to accomplish several tasks within a single action, whereas the time-multiplexed nature of the GUI version forces users to perform one task after the other in a sequential and linear manner. For instance, consider a user who wishes to rotate two blocks in a 90° angle and connect them. In the TUI version of FlowBlocks, this can be accomplished by picking up one block in each hand, rotating them in the appropriate angle and moving them toward each other simultaneously. In contrast, in the GUI version this can only be accomplished by performing a series of separate actions: moving the cursor to the toolbar at the top of the screen, clicking on the Add Block tool, rotating the block to the desired angle, moving the cursor again to the toolbar at the top of the screen, clicking on the Add Block tool, rotating it to the desired angle, and finally moving a selected block toward the other one until they connect.

This inherent difference between the TUI and GUI versions of FlowBlocks also represents the "performance" theoretical theme formulated by Klemmer et al. (2006), which describes how actions are faster and more nuanced compared to symbolic cognition. In light of this theme, coupled with the space-multiplexed nature of TUI, we expected model creation with the TUI version of FlowBlocks to be faster compared to the GUI version.

Another theoretical theme relevant to the FlowBlocks system is Klemmer et al.'s (2006) "visibility" theme. Blocks in the TUI version are more visible than blocks in the GUI version because TUI blocks are actual 3-D objects requiring a relatively large amount of space, whereas GUI blocks are 2-D graphical representations confined within the space of a computer monitor. This difference in visibility is especially salient in regard to unused blocks, because in the GUI version unused blocks are represented by small icons within the toolbar at the top of the screen (see Fig. 2). Since unused blocks are more visible in the TUI version, we expected users to pay greater attention to them, and as a result include more blocks in each model compared to the GUI version.

We believe additional differences regarding performance and preference may exist between the TUI and GUI versions of FlowBlocks, however, in light of the contradicting findings of prior studies comparing TUI to GUI, we found it hard to predict the exact nature of these differences.

2. Method

2.1. Participants

Participants were 58 (39 females, 19 males) undergraduate communications students from an international college in Israel. Their age ranged from 19 to 31 ($M=22.62$, $SD=2.52$). We chose to conduct the experiment with adult college students in order to ensure all our participants have passed a threshold level of cognitive development and computer proficiency. In addition, previous studies comparing GUIs to TUIs involved more child participants than adult participants (see Table 1), so we wished to focus on the less researched population.

While our sample was fairly homogeneous in regard to cognitive development and computer proficiency, participants came from 18 different countries, thus representing a variety of cultures. They were randomly assigned to either GUI ($n=28$) or TUI ($n=30$) condition. All participants volunteered to participate in the study in exchange for extra credit in a class taught by the first author. In order to prevent any bias that might arise from this form of compensation, the first author was not present during the experimental sessions. Furthermore, participants were informed that the same amount of extra credit would be given to everyone who participates, regardless of their performance or expressed opinions during the experimental session.

2.2. The experimental task

The task chosen for the current experiment was unstructured—participants were instructed to freely create models, as many as they can for 30 min, without any additional guidelines or limitations. Such an unstructured task was chosen in order to reflect the original purpose behind FlowBlocks, which is supporting unstructured learning (Zuckerman and Resnick, 2003, 2004, 2005). FlowBlocks were designed in the spirit of the "digital manipulatives" approach (Resnick et al., 1998), stating that students can learn complex concepts through open-ended exploration of computationally-enhanced physical objects.

While the decision to use an unstructured task was rooted in theory, it also has empirical validation. The possible role of task structure was previously evaluated in a study by Marshall et al. (2010), who compared the effects of interface type (GUI vs. TUI) and level of control (low—representing a highly structured task vs. high—representing an unstructured task) on a discovery learning task. They found no significant differences between the low control and high control groups.

After 30 min have passed or participants indicated they have no further ideas (the sooner of the two), they were introduced to the second version of FlowBlocks and were asked to re-create the most complex model they created with the first version. They were allowed 5 min for completing this task. This task asymmetry between the first and second interaction was inevitable. Due to the fact that participants were asked to create models until "they ran out of ideas", we expected the second interaction to be greatly affected by the first interaction. More specifically, we believed that performance and internal experiences would depend on whether or not we allow participants to replicate the same models with both versions of FlowBlocks. If we allow replication, shorter periods of time would be required for creating models with the second version. In addition, participants might get bored. If we do not allow replication, fewer ideas would remain for the second version. In addition, participants might get frustrated. In order to eliminate these potential biases, we opted to have participants experiment extensively with just one version, and then shortly explore the other version to establish a basis for comparison.

Lastly, we opted to have participants perform the experimental task individually, in order to ensure a balanced and standardized comparison between the TUI and GUI versions of FlowBlocks. In regard to balance, TUI has been found to facilitate collaboration among users (Horn et al., 2009). Had our participants worked in groups, the TUI version of FlowBlocks would have had an advantage over the GUI version, thus defeating the purpose of the current experiment. In regard to standardization, various group characteristics have been shown to affect group performance. For instance, group performance was found to be contingent on the distribution of knowledge within the group and networks of social relationships among group members (Rulke and Galaskiewicz, 2000). Had our participants worked in groups, different groups might have been characterized by different levels of knowledge distribution and social cohesiveness. These factors could potentially interact with the effect of the interface, and are beyond the scope of the current study.

2.3. Measures

The study evaluated both actual performance with the GUI and TUI versions of FlowBlocks, as well as potential preference for one version over the other.

2.3.1. Measuring performance

In order to evaluate users' performance while using either the TUI or GUI version of FlowBlocks, video recordings from all experimental sessions were systematically coded by two coders, who remained blind to the research hypotheses. The coders, one male and one female, were undergraduate communications students, who have taken several HCI classes. In three separate sessions, supervised by the authors, the coders learned and practiced the coding guidelines.

In order to establish inter-coder reliability, both coders coded 10 randomly selected sessions (out of 58), half of which were TUI and half GUI. In total, 64 models were coded by both coders. Any disagreements were resolved through discussions with the authors. After several iterations, high levels of reliability have been reached, allowing each coder to continue coding half of the remaining sessions independently. The following coding-based criterions were used to evaluate performance while using the TUI and GUI versions of FlowBlocks (inter-coder reliability indicated in parentheses).

2.3.1.1. Counting-based performance criterions. *Time to create model*—total time required to complete a model ($r_p = .98$, $p < .001$).

No. of blocks in model—number of blocks used to construct a model ($r_p = .99$, $p < .001$).

No. of types of blocks in model—number of types of blocks used to construct a model ($r_p = .97$, $p < .001$).

No. of times model was touched—number of times the user touched (in TUI) or clicked (in GUI) on either the dial on the accumulator block or the handle on the distributer block while creating a model ($r_p = .96$, $p < .001$).

No. of connection errors—number of connection errors encountered by the user while creating a model ($r_p = .94$, $p < .001$). See Section 1.5 for an explanation regarding connection errors.

2.3.1.2. Classification-based performance criterions

Model type—the specific combination of blocks in a model. Five types of models were defined, each representing a different type of “simulation” in the FlowBlocks language: (1) Chain no outflow, (2) Chain with outflow, (3) Chain with distributer, (4) Chain with two accumulators, (5) Closed loop ($Kappa = .95$, $p < .001$). See Section 1.4 for further explanations regarding the various model types.

Model correctness—whether or not a model is considered correct, based on its adherence to the logic of the FlowBlocks language. A model was classified as correct if blocks were connected in a logical way, allowing lights to transfer from one block to another. Otherwise, it was classified as incorrect ($Kappa = 1.0$, $p < .001$).

2.3.2. Measuring preference

Following previous studies (e.g.: Horn et al., 2009; Xie et al., 2008), we evaluated users' preference for either the TUI or GUI version of FlowBlocks by using a questionnaire and an interview, both designed to examine perceived usability—ease of use, intrinsic motivation—affective experiences, and flow—level of immersion and involvement while using each version of FlowBlocks. The questionnaire was filled electronically using the Qualtrics survey system (http://www.qualtrics.com) and included the following scales:

System Usability Scale—Ten items based on Brooke (1996) were used to assess the usability of each version. An example item: “I thought FlowBlocks was easy to use”. All items were rated on a seven-point Likert scale ranging from (1) “strongly disagree” to (7) “strongly agree” (as opposed to the five-point scale originally used by Brooke, in order to achieve a more fine grained conception of participants' experience with the system). Cronbach's alpha reliability for this scale was .75, so a usability index was created based on the mean score of the 10 items ($M = 3.96$, $SD = .80$)[2].

Intrinsic Motivation Inventory—Eighteen items based on McAuley et al. (1989) were used to assess the intrinsic motivation for using each version. An example item: “I enjoyed using FlowBlocks very much”. All items were rated on a seven-point Likert scale ranging from (1) “strongly disagree” to (7) “strongly agree”. Cronbach's alpha reliability for this scale was .79, so an intrinsic motivation index was created based on the mean score of the 18 items ($M = 4.25$, $SD = .72$).

Flow—was assessed with 11 items[3] representing three sub-dimensions of the theoretical concept (Csikszentmihalyi, 1975; Webster et al., 1993). Two items adapted from O'Brien (2010) were used to asses *Control*: “When using FlowBlocks, I felt in control”; “I could not do some of the things I wanted to do with FlowBlocks” (reverse-scored). *Attention Focus* was assessed with three items adapted from Webster et al. (1993): “When using FlowBlocks, I thought about other things” (reverse-scored); “When using FlowBlocks, I was aware of distractions” (reverse-scored); “When using FlowBlocks, I was totally absorbed in what I was doing”; *Cognitive Enjoyment* was assessed with three items adapted from Webster et al. (1993): “Using FlowBlocks was intrinsically interesting”; “Interacting with FlowBlocks excited my

---

[2] The traditional SUS score, based on summing the score contributions from each item, was calculated as well. Both scores yielded the same results, so we opted to use the simpler index based on the mean score of the 10 items.

[3] Since several aspects of flow were already addressed in other scales used in the current experiment, and since several traditional items often used to assess flow were not applicable to FlowBlocks, we chose only the most relevant items from previous studies.

curiosity"; "Using FlowBlocks aroused my imagination"; an item adapted from O'Brien (2010): "I was so involved in FlowBlocks that I lost track of time"; an item adapted from Lee and Tsai (2010): "I have a positive opinion of using FlowBlocks" and an additional original item designed to assess behavioral intention: "I am willing to participate in a similar FlowBlocks study in the future". All items were rated on a seven point Likert scale ranging from (1) "strongly disagree" to (7) "strongly agree". Cronbach's alpha reliability for this scale was .70, so a flow index was created based on the mean score of the 11 items ($M=4.28$, $SD=.82$).

In addition to the questionnaire, a 10-min semi-structured interview was conducted in order to give participants a chance to describe the experience in their own words. They were asked about their overall impression with each version, the advantages and disadvantages of each version, and finally which version they prefer.

All interviews were transcribed and then independently analyzed by two judges—the second author and an undergraduate communications student who participated in several advanced HCI classes. The two judges carefully reviewed the transcriptions and identified emerging common themes. Any disagreements or inconsistencies between the two judges were reconciled by the first author. It is important to note that the thematic analysis of the interviews was performed separately from the video coding described in the previous section and by different people.

2.4. Procedure

Participants were invited to individual sessions in the lab, which were videotaped with their consent. Upon arrival, participants were greeted by a research assistant, who introduced one of the versions, either GUI or TUI (randomly assigned), explained about the purpose of each type of block and demonstrated the creation of a simple three-block model. Once participants understood how the system operates, they were instructed to freely create models, as many as they can. They were asked to indicate the completion of a model by writing a short description or explanation of the model they created. Model creation ended when participants ran out of ideas for models or after 30 min have passed (the earlier of the two). Immediately after experimenting with the system, participants were asked to fill out a questionnaire regarding their experience. Once the questionnaire was filled, the research assistant introduced the participants to the other version, so that participants who previously experimented with the GUI version were now introduced to the TUI version and vice versa. They were asked to use the new version to re-create the most complex model they created earlier. They were allowed 5 min for completing this task. Participants who required less than 5 min were asked to use the remaining time for further exploration of the new version. After participants explored both versions, a short semi-structured interview was conducted by the research assistant. Participants were then debriefed and thanked for their participation.

3. Results

We present our performance and preference results separately, starting with the quantitative analysis, and followed by the qualitative analysis.

3.1. Evaluating performance

Participants were instructed to create as many models as they can, and were allowed up to 30 min for this task. On average, they spent 17:08 minutes actively creating models ($SD=05{:}35$), created 7.45 models ($SD=3.62$) and dedicated 2:18 minutes for each model ($SD=1{:}38$).

In order to compare actual performance between the GUI and TUI versions of FlowBlocks, we preformed a two-way multivariate analysis of covariance (MANCOVA), with version of FlowBlocks (GUI vs. TUI) and gender (male vs. female) as independent variables, the mean score of the counting-based performance criterions as the dependent variables, and the number of error events encountered during model creation as a covariate. This analysis revealed a significant multivariate effect for number of errors, F (4, 424)$=16.58$, $p<.001$; version of FlowBlocks, F (4, 424)$=4.14$, $p<.01$ and gender, F (4, 424)$=2.92$, $p<.05$. However, there was no significant interaction effect between version and gender, F (4, 424)$=.93$, *n.s.* Results of follow-up analyses of variance (ANOVAs), after controlling for the effect of number of errors, are presented in Table 2. As can be seen in the table, participants who used the TUI version of FlowBlocks included more blocks in each model ($M=8.19$, $SD=2.35$) compared to those who used the GUI version ($M=7.42$, $SD=2.59$). Similarly, participants who used the TUI version of FlowBlocks included more types of blocks in each model ($M=5.36$, $SD=.93$) compared to those who used the GUI version ($M=5.08$, $SD=1.08$). In addition, male participants spent more time on each model ($M=02{:}30$, $SD=01{:}46$)

Table 2
Means (standard deviations) and ANOVA results obtained for performance criterions while using the GUI and TUI versions of FlowBlocks, after controlling for the number of errors encountered.

	GUI		TUI		Version		Gender		Version × gender	
	Male	Female	Male	Female	*F*	*p*	*F*	*P*	*F*	*p*
Time to create a model (mm:ss)	02:10 (01:41)	02:07 (01:31)	02:52 (01:48)	02:16 (01:35)	.67	n.s.	4.05	< .05	3.02	n.s.
No. of blocks in model	7.48 (2.52)	7.39 (2.64)	8.42 (2.39)	8.08 (2.34)	4.63	< .05	.70	n.s.	.24	n.s.
No. of types of blocks in model	5.02 (1.08)	5.11 (1.08)	5.36 (.97)	5.37 (.91)	3.95	< .05	.24	n.s.	.18	n.s.
No. of times model was touched	3.46 (3.29)	3.15 (3.97)	6.67 (7.00)	4.70 (5.17)	3.16	n.s.	4.45	< .05	2.84	n.s.

compared to female participants (M=02:11, SD=01:33). Similarly, male participants touched or clicked (depending on the version of FlowBlocks) each model more times (M=4.99, SD=5.60) compared to female participants (M=3.90, SD=4.64). No additional effects were found.

Next, we tested whether the distribution of the classification-based performance criterions differed across version of FlowBlocks and gender. Both classification-based performance criterions – model type and model correctness – were not correlated with the number of errors encountered while creating models.

In regard to model type, an Independent-Samples Mann–Whitney U Test revealed the distribution is the same across version of FlowBlocks ($p > .05$). For both versions of FlowBlocks, the most common type of model was Chain with two accumulators, followed by Chain with distributer, then Chain no outflow, then Closed loop, and lastly Chain with outflow. Specific percentages are presented in Table 3.

Another Independent-Samples Mann–Whitney U Test revealed the distribution of model type differs across gender ($p < .05$). For males, the most common type of model was Chain with two accumulators, followed by Chain no outflow, then Chain with distributer, then Closed loop, and lastly Chain with outflow. For females, the most common type of model was Chain with two accumulators, followed by Chain with distributer, then Chain no outflow, then Closed loop, and lastly Chain with outflow. Specific percentages are presented in Table 3.

In regard to model correctness, Independent-Samples Mann–Whitney U Tests revealed the distribution is the same across version of FlowBlocks as well as across gender (in both cases $p > .05$). Specific percentages are presented in Table 4.

It appears that even though participants were not asked to follow any specific guidelines during the creation of models, the majority of models they created were correct according to the logic of the FlowBlocks language, regardless of the version being used or participants' gender.

In sum, performance was indeed affected by both independent variables tested in the current experimented: version of FlowBlocks and participants' gender, even after controlling for the number of errors encountered while creating models. However, each performance criterion was determined by a different independent variable, with none of the criterions being determined by both variables. More specifically, the number of blocks and the number of types of blocks included in each model was determined only by the version of FlowBlocks—participants tended to include more blocks and more types of blocks in the TUI version. In contrast, the amount of time participants dedicated to each model and the number of times they touched or clicked each model was determined only by gender—males tended to dedicate more time to each model and touch each model more often than females. The distribution of model type differed across gender, but not across version. The distribution of model correctness remained the same across both version and gender.

3.2. Evaluating preference

In addition to actual performance, we were also interested in the subjective experiences while using the two versions of FlowBlocks, and whether participants preferred one version over the other. We evaluated subjective experiences and preference with a questionnaire comprised of three separate scales (usability, intrinsic motivation, flow) and an interview.

3.2.1. Questionnaire analysis

In order to compare participants' subjective experiences while using the GUI and TUI versions of FlowBlocks, we preformed a two-way univariate analysis of covariance (ANCOVA) for each scale of the questionnaire, with version of FlowBlocks (GUI vs. TUI) and gender (male vs. female) as independent variables, and the number of error events encountered during model creation as a covariate. However, unlike the performance criterions, none of the scales in the questionnaire were affected by the number of errors. We therefore dropped this variable from further analysis.

Table 4
Distribution of model correctness across version of FlowBlocks and gender.

	Model correctness	
	Correct (%)	Incorrect (%)
Version		
TUI	81.6	18.4
GUI	78.7	21.3
Gender		
Male	77.9	22.1
Female	81.2	18.8

Table 3
Distribution of model type across version of FlowBlocks and gender.

	Model type				
	Chain no outflow (%)	Chain with outflow (%)	Chain with distributer (%)	Chain with two accumulators (%)	Closed loop (%)
Version					
TUI	10.6	2.9	17.9	64.3	4.3
GUI	16.9	3.1	20.4	48.4	11.1
Gender					
Male	13.1	2.1	11.7	62.8	10.3
Female	14.3	3.5	23.0	52.6	6.6

Table 5 shows means, standard deviations and ANOVA results obtained for the System Usability, Intrinsic Motivation and Flow scales.

In regard to *Usability*, there was a significant main effect for version. There was no significant main effect for gender, but there was a significant interaction between version and gender. In order to better understand the nature of this interaction, two simple ANOVAs were conducted separately for males and females. Males rated the usability of the GUI version of FlowBlocks significantly higher than the usability of the TUI version, $F(1, 17) = 13.92$, $p < .01$. In contrast, females did not show such a preference and rated the usability of the GUI version and the TUI version similarly, $F(1, 37) = 1.36$, *n.s.*

In regard to *Intrinsic Motivation*, there was no main effect for either version or gender, and no interaction effect between these two factors. It seems that both versions are equally motivating for both males and females. Similarly, in regard to *Flow*, there was no main effect for either version or gender, and no interaction effect between these two factors. It seems that both versions are equally engaging for both males and females.

A correlation matrix between performance criterions and subjective experience scales is presented in Table 6. Interestingly, the performance criterions are correlated to one another, and the subjective experience scales are mostly correlated to one another, but there are almost no correlations between these two groups. The only significant correlation between a performance criterion and a subjective experience scale is a correlation between the number of times a model was touched or clicked and flow. The number of times a model was touched or clicked represents the amount of simulation the participant performed with this model, so that more touches reflect more simulation. Therefore, it is not surprising that a participant who performed more simulation, in other words "played more with the system", experienced a greater sense of flow.

3.2.2. *Interview analysis*

In addition to the quantitative scales, we enabled our participants to express their subjective experiences with both versions of FlowBlocks in their own words, through a semi-structured interview. During the interview they were asked about their overall impression with each version, the advantages and disadvantages of each version, and lastly which version they prefer. All interviews were transcribed and then independently analyzed by two judges, who carefully reviewed the transcriptions and identified emerging common themes. Results of this thematic analysis are presented below, with quotes from participants' comments to demonstrate each theme.

3.2.2.1. Overall impression. First, participants were asked about their overall impression with each version, focusing mainly on the valence of the interaction and whether it was easy to learn. Most participants described the interaction with both the TUI and GUI versions of FlowBlocks as a positive experience, referring specifically to how enjoyable and interesting it was. In fact, the most commonly used adjectives to describe the experience with both versions were "fun" and "interesting".

> TUI: "It was fun to play with the notion of flow, and see how I can control the flow".
>
> GUI: "It might come off as something silly, but it's actually something pretty helpful and fun".

Table 6
Correlation matrix between performance criterions and subjective experience scales.

	Time to create model	No. of blocks in model	No. of types of blocks in model	No. of times model was touched	System usability	Intrinsic motivation	Flow
Time to create model	1.0						
No. of blocks in model	.63**	1.0					
No. of types of blocks in model	.56**	.90**	1.0				
No. of times model was touched	.65**	.32*	.35**	1.0			
System usability	−.05	−.11	−.06	.02	1.0		
Intrinsic motivation	.05	−.01	.08	.18	.45**	1.0	
Flow	.25	.20	.26	.27*	.19	.65**	1.0

*Correlation is significant at the $p < .05$ level.
**Correlation is significant at the $p < .01$ level.

Table 5
Means (standard deviations) and ANOVAs results obtained for usability, intrinsic motivation and flow ratings.

	GUI		TUI		Version		Gender		Version x gender	
	Male	Female	Male	Female	*F*	*p*	*F*	*p*	*F*	*p*
Usability	4.59 (.37)	4.11 (.78)	3.42 (.87)	3.82 (.75)	12.34	.001	.41	n.s.	4.56	.05
Intrinsic motivation	4.35 (.39)	4.05 (.64)	4.34 (.77)	4.36 (.87)	.61	n.s.	.47	n.s.	.62	n.s.
Flow	4.35 (.73)	4.03 (.87)	4.55 (.75)	4.35 (.84)	1.27	n.s.	1.26	n.s.	.09	n.s.

Though both versions elicited positive reactions, the TUI version was perceived as more playful than the GUI version, often bringing up childhood memories.

> TUI: “I had blocks when I was a kid, so I definitely like blocks, and these blocks have lights, which makes them super fun”.

A negative experience with one of the versions was usually the result of extraneous factors unrelated to the system itself, for instance feeling uncomfortable in a lab setting.

> TUI: “I think it was only negative because I was being recorded. I think if I didn't know there is a camera there, it would have been nice”.

Several participants explained that they had a negative experience due to technical problems or inability to complete a task.

> TUI: “My impression was very poor because you have to disconnect the wires, and the wire… it just didn't flow”.
> GUI: “It was a little bit frustrating because sometimes you can't do what you want to do, and you have to deal with the way the blocks were made before, you can't change them, so it's a little bit frustrating”.

Most participants easily understood how to operate each version, requiring only a small number of trials before properly using both systems.

> TUI: “Not hard at all [to learn] because it's all magnetic, so it was real easy to attach everything, you can see what side is magnetic to the other because the pins come out, and the others are inserted, they are just easy to handle”.
> GUI: “Not hard at all [to learn] cause it's just click, drag, click. Any person that uses a computer can function very well”.

In sum, it seems that both versions elicited mostly positive reactions and were enjoyable and easy to learn. The TUI version was perceived as more playful or game-like compared to the GUI version.

3.2.2.2. Advantages and disadvantages. Participants were asked which advantages and disadvantages they saw in each version. The most commonly mentioned advantage of the TUI version was its tangibility—the ability to manipulate it by hand and receive rich feedback. These properties enhanced participants' sense of realism and control over the interaction.

> “It's a little more natural to be working with [actual blocks], it's more hands-on, it's in front of you, you're doing something, you're touching something, it's more sensory”.

Several participants also mentioned that it was easier for them to understand how to interact with the system while using the TUI version.

> “It was easier to kind of get an idea of how everything worked, and where the power was coming from, and why you need to put certain blocks and where”.

The most commonly mentioned disadvantage of the TUI version was the fact that it was prone to technical errors. The physical connections between the blocks are magnetic, and sometimes require users to disconnect and reconnect them again in order to assure a continuous flow of light.

> “The connections were not that good, you had to mess with them a few times to get them to work”.

The TUI version was also perceived as cumbersome, as it is not easily transported from one location to another and requires a large work space.

> “It's big and bulky so it's not so easy to pick up and move around”.

In addition, several participants mentioned that model creation was slower in the TUI version compared to the GUI version.

> “Takes longer to assimilate and disconnect”.

Lastly, the blocks themselves were perceived as potentially fragile.

> “I was afraid it would break or something… It seems a little bit loose”.

The most commonly mentioned advantage of the GUI version was its reliability.

> “On the computer you didn't have so many issues with the connection points”.

The second most mentioned advantage of the GUI version was its efficiency.

> “It takes less time, cause you can put things back with the click of a mouse”.

Another commonly mentioned advantage of the GUI version was the flexible view, which was easily adjusted with “zoom in” and “zoom out” buttons, and also counted the remaining blocks of each type.

> “You could zoom in and zoom out, it was really good, you could focus on different things”.

Theoretically, a similar “functionality” can be achieved in the TUI version by physically moving closer to or away from the table and manually grouping blocks together according to type, but that requires more effort.

Several participants mentioned that the GUI version required less effort.

> “The computer attaches blocks for you, and [in TUI] you have to do it by yourself”.

Another perceived advantage of the GUI version was the fact that it is portable and manageable.

> “You don't actually physically need to have the blocks, you can download it from anywhere”.

The most commonly mentioned disadvantage of the GUI version was the fact that it was digital, and therefore perceived as less natural or real.

> “It can be annoying, like twisting and turning it using the mouse, it doesn't feel natural”.

In sum, the TUI version of FlowBlocks provides physical stimulation that enhances perceived realism and control, but it is prone to technical errors, which make it less reliable and less efficient. It is also less portable compared to the GUI version. In contrast, the GUI version of FlowBlocks is reliable, efficient, adjustable and effortless, but perceived as less natural and less real.

3.2.2.3. Preference. Lastly, participants were asked which version they prefer—GUI or TUI. Preference rates are shown in Table 7. The majority of participants preferred the TUI (62.1%) over the GUI (34.5%) version, and only 2 participants (3.4%) were unable to state a preference. A one-sample binominal test revealed that the preference for TUI is significant at the $p < .05$ level. Furthermore, this preference was unrelated to gender ($r_c = .05$, *n.s.*), order of using the two versions ($r_c = .08$, *n.s.*) or participants' country of origin ($r_c = .48$, *n.s*).

This preference is surprising, considering that the GUI version was perceived as more reliable and efficient than the TUI version, it received higher usability ratings from men, and no significant differences between the two versions were observed in regard to any of the other subjective measures. So why did participants prefer the apparently inferior version? It seems that they were willing to overlook the lower reliability and efficiency because the TUI version was more enjoyable and stimulating.

> “You might have technological problems, but you can connect them with your hands, and not with a mouse click, you can turn them around, you can put them together, and feel you can put them together. You feel the magnetism you know…so this is much more of an experience”.
> ”[GUI] is easier, it's always working, I don't have to check the… tiny technical stuff, but [TUI] is more of a game”.

According to most participants, the TUI version is “more fun” because it provides rich feedback, enables physical interaction (feeling, not only seeing and hearing), and produces high levels of realism and interactivity.

> “I like the real thing better, it's a better feeling when you connect…and it does the “click”, and seeing the light, and touching the [dial]. In the real thing it feels better than in the computer”.

Table 7
Preference rates for TUI and GUI versions of FlowBlocks.

	Prefer GUI (%)	Prefer TUI	No preference
Interacted with GUI first	32.1	67.9	.0
Interacted with TUI first	36.7	56.7	6.7

> “I think [the blocks] had more impact because you can actually see what you are making. Like in the real world, not in a virtual world. I'm more of a real person than a virtual person”.

The TUI version is also perceived as encouraging creativity.

> “People would be more creative having the actual blocks in their hands”.

In addition, some participants preferred the TUI version over the GUI one simply because they prefer not to interact with computers.

> “I prefer [TUI] version because I don't like computers”.

Preference for GUI over TUI usually derived from its superior efficiency and reliability.

> “On the screen you can see everything, you have better view of all the blocks, and it's handier, because it's easy to drag it, and attach it, it's not heavy like the real blocks”.

Several participants explained that they prefer the GUI version because the different elements of the system were clearer in that version.

> “On the computer you can see easier, like it's clear, it's more clear than the blocks. Also to see the light flows, and if you put a block here that doesn't go with it, that doesn't work, you can see it better, you can see the numbers, and you can change them better—the positions, and the blocks”.

In addition, some participants preferred the GUI version over the TUI one simply because they prefer to interact with computers.

> “Everything is on a computer now. I just like the computer better. The digital feels better”.

Only few participants said they prefer the GUI version because “it is more fun to play with”.

A summary of the main attributes of each version of FlowBlocks appears in Table 8.

Most participants explained that they prefer the TUI version over the GUI version due to its tangibility or “hands-on” nature. Since we measured the number of times each model was actually touched (in TUI) or clicked on (in GUI), we tested whether preference is correlated with the number of touches and clicks, using an independent-samples *t*-test. Indeed, there was a significant difference in mean number of touches and clicks per model between the two versions of FlowBlocks, $t(50) = -2.33$, $p < .05$. Participants who ended up preferring the TUI version touched and clicked each model more ($M = 2.90$, $SD = 2.48$) compared to participants who ended up preferring the GUI version ($M = 1.82$, $SD = .93$). No additional differences were found between those who preferred the TUI version and those who preferred the GUI version in regard to any other performance criterion or subjective experience scale, not even number of errors encountered. These findings provide further support to the notion that the TUI version is preferred over the GUI version mainly due to its tangibility.

Table 8
Attributes of the TUI and GUI versions of FlowBlocks, as reflected in interviews with users.

Attribute	TUI	GUI
Enjoyment	Very high	High
Playfulness	High	Low
Ease of learning	High	High
Reliability	Low	High
Efficiency	Low	High
Effort	High	Low
Physical stimulation	High	Low
Realism	High	Low
Portability	Low	High

4. Discussion

Tangible user interfaces are assumed by some to have added advantages compared to graphical user interfaces (Fitzmaurice and Buxton, 1997; Fitzmaurice et al., 1995; Zuckerman et al., 2005), yet several empirical studies were unable to demonstrate clear advantages for either interface (e.g. Horn et al., 2009; Marshall et al., 2010), and others yielded contradicting findings. As a result, the relative strengths and weaknesses of each interface remain unclear. The current study set out to explore these strengths and weaknesses, by comparing between the GUI and TUI versions of the same system, called FlowBlocks. While developing the two versions, we aimed to strike a balance between making them as similar as possible to ensure an accurate comparison, and being careful not to "control out" all the unique affordances of each interface. While designing the study, we aimed for a controlled and unbiased environment, and included multiple evaluative measures to ensure a comprehensive comparison.

The results of the study revealed that most participants preferred the TUI version over the GUI version. This preference was statistically significant and characterized both men and women, regardless of which version they were introduced to first or interacted with longer. This preference for the TUI version is surprising, because in most other aspects the two versions were equivalent, and in some aspects the TUI version was even inferior to the GUI version.

In regard to performance parameters, due to the bimanualism and space-multiplexed nature of TUI vs. the time-multiplexed nature of GUI (Fitzmaurice and Buxton, 1997), we expected model creation with the TUI version of FlowBlocks to be faster compared to the GUI version. In contrast to our expectation, the amount of time required to create models was similar for both versions. However, one must remember that the TUI version was more prone to connectivity errors compared to the GUI version. Therefore, while a similar amount of time was required in both versions, participants using the TUI version had to both create models and engage in troubleshooting, whereas participants using the GUI version only created models. Consequently, one possible explanation is that model creation was indeed faster with the TUI version, but this advantage was masked by the errors with which participants had to contend.

Due to the higher visibility of TUI (Klemmer et al., 2006), especially in regard to unused blocks, we expected users to include more blocks in models created with the TUI version compared to the GUI version. As expected, participants who used the TUI version included more blocks and more types of blocks in each model. In addition to heightened visibility, it is possible the inclusion of additional blocks in the TUI version reflects a more inviting nature of the interface (Horn et al., 2009), encouraging participants to interact with it and explore the various types of available blocks.

In regard to other performance criterions in our study, the two versions of FlowBlocks were equivalent: the distribution of model type was similar in both versions, as was the frequency of correct vs. incorrect models. In both the GUI and TUI versions, the majority of models (approximately 80%) were correct according to the logic of the FlowBlocks language. These findings suggest that participants were able to successfully learn FlowBlocks even without receiving structured instructions.

In regard to questionnaire ratings, flow and intrinsic motivation were rated similarly for both versions of FlowBlocks. However, the TUI version received lower usability ratings from men and was generally perceived as less reliable and efficient than the GUI version.

Interestingly, the questionnaire ratings were mostly correlated to one another, and the various performance criterions were also correlated to one another, but there were almost no correlations between these two groups of metrics. This finding implies that performance and preference are two independent aspects of the user experience. Indeed, the TUI version was preferred over the GUI version despite its lower usability and without exhibiting clear advantages in regard to performance.

Why did participants prefer the apparently inferior interface? Insights from interviews with participants provided us with the answer. It seems our participants were willing to overlook the low usability of the TUI version simply because it was more stimulating and enjoyable compared to the GUI version. Which properties of the TUI version made it more stimulating and enjoyable? First, it enables physical interaction, as evident in participants' statements like: "I was more physically involved" and "The hands-on is a big thing". Second, it provides rich feedback, as evident in statements like: "It's a better feeling when you connect and it does the 'click', and seeing the light, and touching the dial". Third, it produces high levels of realism, as evident in statements like: "It is more interesting to play with something real than to play with the flat screen". In other words, interaction with the TUI version of FlowBlocks is perceived as a "real" experience, whereas interaction with the GUI version is perceived as a "virtual" experience (even though both versions are digital), and as one participant explained: "I'm more of a real person than a virtual person".

If interacting with the TUI version is indeed a more enjoyable experience than interacting with the GUI version, a more interactive use of the system is expected to strengthen the preference for the TUI version. Our quantitative analysis revealed exactly that—participants who interacted more with the system (touched the dials and distributer handles in the TUI version or clicked on them in the GUI version) tended to prefer the TUI version over the GUI version more than participants who interacted less with the system.

Why did physical interaction, rich feedback, and high levels of realism emerge as reasons for preferring TUI? We suggest these properties characterize intuitive interactions with the environment. One of the most intuitive forms of interaction with the physical world is picking up a selected object, manipulating it with our hands, and receiving feedback based on our actions. We are accustomed to and highly skilled in this form of interaction. The three properties of TUI mentioned by our participants are exactly those properties that make interactions with TUI resemble intuitive interactions.

Thinking more broadly, our findings could also be explained through Norman's theory of beauty. According to Norman (2004), there are three levels of processing beauty: a visceral level, referring to surface beauty; a behavioral level, referring to beauty in operation; and a reflective level, referring to beauty in meaning and implication. We believe tangible interactions are processed at the behavioral level, defined by Norman as “the home of highly learned skills”. At this level, positive affect stems from feeling in control and understanding during the use of a product. Behavioral responses like these play an important role in subjective experiences. According to this view, preference for the TUI version of FlowBlocks stemmed from participants' prior experience with tangible interactions, which helped them understand how to operate the system and feel confident in their actions. Indeed, the interaction with FlowBlocks often aroused childhood memories of playing with blocks. Moreover, Norman believes that higher levels of processing could override lower ones. In our case, the positive subjective experiences with TUI were strong enough to override the objectively lower usability.

Additional interesting findings were obtained in the current study in regard to gender differences. Men spent more time and interacted with each model more than women. Men also rated the usability of the GUI version higher than that of the TUI version, while women's usability ratings did not differ between the two versions. Gender differences in responses to TUI and GUI systems in particular, and in regard to technology acceptance in general, were found in previous studies as well. For example, Horn et al. (2009) found that museum visitors were more likely to try an exhibit when it involved tangible blocks rather than a mouse, and that this effect was stronger for girls than it was for boys. Similarly, Xie et al. (2008) found that ratings of interest and enjoyment were nearly the same for all gender pairings when using TUI, but differed when using GUI, so that boy–boy pairs gave significantly higher ratings than the girl–girl and girl–boy pairs. More generally, men and women were shown to differ in the salience of various factors determining technology adoption and usage behavior, so that men consider perceived usefulness to a greater extent than women (Venkatesh and Morris, 2000).

In sum, the main findings of the current study are in line with those of several previous studies, which were unable to demonstrate clear performance advantages for either graphical or tangible interfaces (e.g.: Horn et al., 2009; Marshall et al., 2010). However, unlike previous studies, we were able to demonstrate a clear preference for the tangible interface, stemming from favoring stimulation and enjoyment (TUI) over usability (GUI). Three main characteristics of the TUI version make it more stimulating and enjoyable: it enables physical interaction, it provides rich feedback, and it produces high levels of realism.

While the preference for TUI was clearly evident in our interview analysis, it was not reflected in the questionnaire analysis. This incongruence between the quantitative and qualitative findings deserves further discussion. We employed quantitative questionnaires and qualitative interviews to measure similar concepts—perceived usability, intrinsic motivation and sense of flow. Interview analysis revealed clear differences in regard to all three, whereas questionnaire analysis revealed significant differences only in regard to usability. Questionnaire-based findings regarding motivation and flow were similar to those revealed in the interviews – TUI received slightly higher ratings than GUI – but the difference did not reach statistical significance. We believe the reason behind this incongruence is the different type of analysis performed on questionnaire-based data and interview-based data. As explained in Section 2, in order to eliminate potential biases we opted to have participants fill out a questionnaire regarding only one version of FlowBlocks, either TUI or GUI, as determined by random assignment. Consequently, a between-subjects analysis was performed on the questionnaire data (the different versions of FlowBlocks were evaluated by comparing the ratings of two separate groups of participants). In contrast, interviews were conducted after participants interacted with both versions of FlowBlocks and were able to conduct a direct comparison between the two. Consequently, a within-subjects perspective was adopted when analyzing interview data (the different versions of FlowBlocks were evaluated within the same participant). Within-subjects analysis is more sensitive, therefore we are not surprised it was able to reveal differences that the between-subjects analysis could not. Furthermore, the various attributes associated with the TUI and GUI versions of FlowBlocks, as summarized in Table 8, are relative in nature. It is plausible that our participants found it hard to judge each version on its own merit, so when filling out the questionnaire they leaned towards the mid-point of the scale (mean ratings were approximately 4, on a 7-point scale). Only after being introduced to the second version could they formulate more extreme judgments based on a direct comparison, and express them during the interview. If this is indeed the case, future studies designed to compare between different interfaces should adopt a within-subjects approach.

The present study includes several limitations that should be addressed. First, we compared between the GUI and TUI versions of a single prototype—FlowBlocks. The preference we found for a stimulating tangible interface over an efficient graphical interface should be demonstrated in regard to additional systems.

Second, as explained in Section 2, in order to eliminate potential biases we opted to have participants experiment extensively with one version of FlowBlocks (either TUI or

GUI, randomly assigned), and then only briefly explore the other version (GUI or TUI, respectively) to establish a basis for comparison. This asymmetry could potentially lead participants to prefer the more familiar version, as suggested by the "mere exposure effect" (Zajonc, 1968). According to this well-documented effect, people tend to develop a preference for things merely because they are familiar with them. Alternatively, a novelty effect (Cantor, 1968) could have been operating, leading participants to prefer the less familiar version. If a mere exposure effect had played a role here, participants should have preferred the first version with which they interacted, because this interaction was longer. In contrast, if a novelty effect had played a role here, participants should have preferred the second version with which they interacted, because the interaction was shorter. Our findings indicated that order did not play a significant role—both participants in the GUI-first and the TUI-first conditions preferred the TUI version over the GUI version. Furthermore, despite having 30 min for experimenting with the first version of FlowBlocks, participants utilized only 17 min on average, suggesting that at least part of the novelty subsided even within the experimental session. Still, we cannot completely disprove a novelty effect, as GUIs are generally more common than TUIs, and interactions in the lab could never measure up to participants' life-long experiences.

Third, the current study focused only on adult population, because our literature review indicated it was less explored in the context of TUI compared to child population. It would be interesting to compare preference rates for TUI and GUI between different age groups. If the preference for TUI indeed reflects a yearning for natural interactions in a growingly digital world, the preference might be more evident among adults and teenagers, who typically engage in digital interactions more frequently than young children.

An additional limitation stems from the short-term nature of the current evaluation. The study was based on a single session, therefore we cannot simply assume that the preference for TUI over GUI would persist in the long-term as well. Prior research demonstrated that the effect of hedonic factors on intentions to explore a technology decreases over time, while the effect of instrumental factors increases (Magni et al., 2010). If the same happens with FlowBlocks, a reversal of preference would occur, turning the GUI version into the more preferable one. Future studies should examine such long-term effects.

Lastly, the current study included only a single comparison between GUI and TUI conditions, but several other relevant comparisons could be made. For instance, adding a control condition in which participants are asked to create models with non-digital materials, like paper cards or simple wooden blocks, could serve as a baseline for evaluating the benefits of using any digital system. Along the same lines, it would be interesting to compare TUI and GUI with the multitouch interface of tablet computers and smartphones. Similar to TUI, multitouch interfaces enable physical interaction, however, unlike TUI, they do not offer rich feedback or realism because the user only touches a flat screen. In that regard, multitouch interfaces are limited, because they cannot fully replicate natural interactions. Therefore, we would expect preference rates for multitouch interfaces to fall between those for TUI and GUI.

Despite the above mentioned limitations, the current findings have several practical implications. If tangible interfaces are perceived as more stimulating and enjoyable, they can be utilized to enhance excitement regarding tasks that would otherwise not be perceived as particularly exciting. For example, educators could introduce TUI-based systems in order to strengthen students' engagement in specific domains, and managers could introduce TUI-based systems in order to make workers' tedious or repetitive tasks seem more stimulating, at least in the short term. However, when the task at hand requires efficiency, a GUI-based system might be preferable to a TUI-based system. It seems worthwhile to explore this "interface-task fit" hypothesis in future studies.

Another possibility, suggested by Horn et al. (2012), is combining tangible and graphical interfaces within the same system. This hybrid approach offers users the flexibility to select the most suitable interface for their needs, and to easily switch to another interface should a new need arise.

More generally, the findings of the current study highlight the importance of the subjective experience with a system, which was proven powerful enough to override even objective properties. Norman (2004) claimed that "we must move beyond products that perform well toward products that deliver pleasure, beauty, and fun" (p. 317). Our study revealed one way for achieving this goal—introducing TUI-based systems.

Whether designing hybrid or separate interfaces, HCI researchers and practitioners should consider the strengths and weaknesses of tangible and graphical interfaces, and select the one most suitable for their specific needs.

5. Conclusions

Tangible user interfaces are assumed to have added advantages compared to graphical user interfaces. However, due to contradicting findings from different empirical studies, this assumption cannot be supported by the existing literature. The current study set out to further our understanding of the relative strengths and weaknesses of each interface, by comparing between similar GUI and TUI versions of the same system—a modeling and simulation tool called "FlowBlocks". Most participants preferred the TUI version over the GUI version. This preference was statistically significant and characterized both men and women, regardless of which version they were introduced to first or interacted with longer. This preference for the TUI version over the GUI version is surprising, because the two versions were equivalent in regard to most performance parameters. Furthermore, the TUI version was perceived as inferior compared to the GUI version in regard to usability. Interviews with users revealed that the TUI version was preferred over the GUI version because it (1) enables physical interaction (2) provides rich feedback (3) produces high levels of realism—properties that make the TUI version stimulating and enjoyable. These findings highlight the importance of the subjective experience with a system, which can be powerful enough to override even objective properties.

Acknowledgments

We would like to thank Mr. Shlomi David and Ms. Meytal Abo for their assistance with data collection; Mr. Ran Frajman for his assistance with development of the GUI version of FlowBlocks; Mr. Itamar Gilon and Ms. Sharon Heller for their assistance with video analysis; Ms. Maya Hofri for her assistance with interview analysis; Dr. Doron Friedman, Dr. Guy Hoffman, Dr. Winslow Burleson and three anonymous reviewers for their insightful comments on early versions of this paper.

References

Booth Sweeney, L., Sterman, J.D., 2000. Bathtub dynamics: initial results of a system thinking inventory. System Dynamics Review 16, 249–286.

Broaders, S., Cook, S.W., Mitchell, Z., Goldin-Meadow, S., 2007. Making children gesture brings out implicit knowledge and leads to learning. Journal of Experimental Psychology: General 136, 539–550.

Brooke, J., 1996. SUS: a quick and dirty usability scale. In: Jordan, P.W., Thomas, B., Weerdmeester, B.A., McClelland, I.L. (Eds.), Usability Evaluation in Industry. Taylor & Francis, Londonpp. 189–194.

Buechley, L., Eisenberg, M., 2007. Boda blocks: a collaborative tool for tangible three-dimensional cellular automata. In: Proceedings of CSCL'07, pp. 102–104.

Cantor, G.N., 1968. Children's "like-dislike" ratings of familiarized and nonfamiliarized visual stimuli. Journal of Experimental Child Psychology 6, 651–657.

Cheng, L.K., Der, C.S., Sidhu, M.S., Omar, R., 2011. GUI vs. TUI: engagement for children with no prior computing experience. Electronic Journal of Computer Science and Information Technology 3, 31–39.

Csikszentmihalyi, M., 1975. Beyond Boredom and Anxiety. Jossey-Bass, San Francisco.

Dourish, P., 2001. Where the Action is: the Foundations of Embodied Interaction. MIT Press, Cambridge, MA.

Fitzmaurice, G.W., Buxton, W., 1997. An empirical evaluation of graspable user interfaces: towards specialized, space multiplexed input. In: Proceedings of CHI'97, pp. 43–50.

Fitzmaurice, G.W., Ishii, H., Buxton, W., 1995. Bricks: laying the foundations for graspable user interfaces. In: Proceedings of CHI'95, pp. 442–449.

Goldin-Meadow, S., Cook, S.W., Mitchell, Z.A., 2009. Gesturing gives children new ideas about math. Psychological Science 20, 267–272.

Hecht, D. Reiner, M., Halevy, G., 2005. Multi-modal stimulation, response time and presence. In: Proceedings of PRESENCE'05, pp. 269–274.

Hecht, D., Reiner, M., Karni, A., 2008. Enhancement of response times to bi- and tri-modal sensory stimuli during active movements. Experimental Brain Research 185, 655–665.

Hmelo, C.E., Holton, D.L., Kolodner, J.L., 2000. Designing to learn about complex systems. Journal of the Learning Sciences 9, 247–298.

Hmelo-Silver, C.E., Pfeffer, M.G., Malhotra, B.A., 2003, April. Fish swim and rocks sit: understanding structures, behaviors, and functions in a complex system. Paper Presented at the American Educational Research Association Annual Meeting, Chicago, IL.

Horn, M.S., 2008. Tangible computer programming for informal science learning. In: Proceedings of IDC'08, pp. 21–24.

Horn, M.S., Crouser, R.J., Bers, M.U., 2012. Tangible interaction and learning: the case for a hybrid approach. Personal and Ubiquitous Computing 16, 379–389.

Horn, M.S., Solovey, E.T., Crouser, R.J., Jacob, R.J.K., 2009. Comparing the use of tangible and graphical programming languages for informal science education. In: Proceedings of CHI'09, pp. 975–984.

Hornecker, E., Buur, J., 2006. Getting a grip on tangible interaction: a framework on physical space and social interaction. In: Proceedings of CHI'06, pp. 437–446.

Ishii, H., Ullmer, B., 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In: Proceedings of CHI'97, pp. 234–241.

Jacob, R.J.K., Ishii, H., Pangaro, G., Patten, J., 2002. A tangible interface for organizing information using a grid. In: Proceedings of CHI'02, pp. 339–346.

Klemmer, S.R., Hartmann, B., Takayama, L., 2006. How bodies matter: five themes for interaction design. In: Proceedings of DIS'06, pp. 140–149.

Lee, M.C., Tsai, T.R., 2010. What drives people to continue to play online games? An extension of technology model and theory of planned behavior. International Journal of Human-Computer Interaction 26, 601–620.

Magni, M., Taylor, M.S., Venkatesh, V., 2010. 'To play or not to play': a cross-temporal investigation using hedonic and instrumental perspectives to explain user intentions to explore a technology. International Journal of Human-Computer Studies 68, 572–588.

Manches, A., O'Malley, C., Benford, S., 2009. Physical manipulation: evaluating the potential for tangible designs. In: Proceedings of TEI'09, pp. 77–84.

Marshall, P., 2007. Do tangible interfaces enhance learning? In: Proceedings of TEI'07, pp. 163–170.

Marshall, P., Cheng, P.C.-H., Luckin, R., 2010. Tangibles in the balance: a discovery learning task with physical or graphical materials. In: Proceedings of TEI'10, pp. 153–160.

Marshall, P., Fleck, R., Harris, A., Rick, J., Hornecker, E., Rogers, Y., Yuill, N., Dalton, N.S., 2009. Fighting for control: children's embodied interactions when using physical and digital representations. In: Proceedings of CHI'09, pp. 2149–2152.

McAuley, E., Duncan, T., Tammen, V.V., 1989. Psychometric properties of the intrinsic motivation inventory in a competitive sport setting: a confirmatory factor analysis. Research Quarterly for Exercise and Sport 60, 48–58.

Newton-Dunn, H., Nakano, H., Gibson J., 2003. Block Jam: a tangible interface for interactive music. In: Proceedings of NIME'03, pp. 170–177.

Norman, D.A., 2004. Introduction to this special section on beauty, goodness, and usability. Human-Computer Interaction 19, 311–318.

O'Brien, H.L., 2010. The influence of hedonic and utilitarian motivations on user engagement: the case of online shopping experiences. Interacting with Computers 22, 344–352.

Patten, J., Ishii, H., 2000. A comparison of spatial organization strategies in graphical and tangible user interfaces. In: Proceedings of DARE'00, pp. 41–50.

Raffle, H.S., Parkes, A.J., Ishii, H., 2004. Topobo: a constructive assembly system with kinetic memory. In: Proceedings of CHI'04, pp. 647–654.

Resnick, M., Martin, F., Berg, R., Borovoy, R., Colella, V., Kramer, K., Silverman, B., 1998. Digital manipulatives: new toys to think with. In: Proceedings of CHI'98, pp. 281–287.

Rogers, Y., Scaife, M., Gabrielli, S., Smith, I.I., Harris, E.A., 2002. Conceptual framework for mixed reality environments: designing novel learning activities for young children. Presence: Teleoperators and Virtual Environments 11, 677–686.

Rulke, D.L., Galaskiewicz, J., 2000. Distribution of knowledge, group network structure, and group performance. Management Science 46, 612–625.

Senge, P.M., 1990. The Fifth Discipline: The Art and Practice of the Learning Organization. DoubleDay, New York.

Shaer, O., Hornecker, E., 2010. Tangible user interfaces: past, present, and future directions. Foundations and Trends in Human–Computer Interaction 3, 1–137.

Shaer, O., Jacob, R.J.K., 2009. A specification paradigm for the design and implementation of tangible user interfaces. ACM Transactions on Computer–Human Interaction 16, 1–39.

Sterman, J.D., 2000. Business Dynamics: Systems Thinking and Modeling for a Complex World. McGraw-Hill, Boston.

Stevanoni, E., Salmon, K., 2005. Giving memory a hand: Instructing children to gesture enhances their event recall. Journal of Nonverbal Behavior 29, 217–233.

Suzuki, H., Kato, H., 1995. Interaction-level support for collaborative learning: AlgoBlock—an open programming language. In: Proceedings of CSCL'95, pp. 349–355.

Terrenghi, L., Kirk, D., Sellen, A., Izadi, S., 2007. Affordances for manipulation of physical versus digital media on interactive surfaces. In: Proceedings of CHI'07, pp. 1157–1166.

Tuddenham, P., Kirk, D., Izadi, S., 2010. Graspables revisited: multi-touch vs. tangible input for tabletop displays in acquisition and manipulation tasks. In: Proceedings of CHI'10, pp. 2223–2232.

Ullmer, B., Ishii, H., 1999. Mediablocks: tangible interfaces for online media. In: Proceedings of CHI'99, pp. 31–32.

Ullmer, B., Ishii, H., 2000. Emerging frameworks for tangible user interfaces. IBM Systems Journal 39, 915–931.

Venkatesh, V., Morris, M.G., 2000. Why don't men ever stop to ask for directions? Gender, soial influence, and their role in technology acceptance and usage behavior. MIS Quarterly 24, 115–139.

Webster, J., Trevino, L.K., Ryan, L., 1993. The dimensionality and correlates of flow in human–computer interactions. Computers in Human Behavior 9, 411–426.

Wyeth, P., Purchase, H.C., 2002. Tangible programming elements for young children. In: Proceedings of CHI'02, pp. 774–775.

Xie, L., Antle, A.N., Motamedi, N., 2008. Are tangibles more fun? Comparing children's enjoyment and engagement using physical, graphical and tangible user interfaces. In: Proceedings of TEI'08, pp. 191–198.

Zajonc, R.B., 1968. Attitudinal effects of mere exposure. Journal of Personality and Social Psychology 9, 1–27.

Zuckerman O., Arida S., Resnick M. 2005. Extending tangible interfaces for education: digital Montessori-inspired manipulatives. In: Proceedings CHI'05, pp. 859–868.

Zuckerman, O., Grotzer, T., Leahy, K., 2006. Flow blocks as a conceptual bridge between understanding the structure and behavior of a complex causal system. In: Proceedings of the 7th International Conference on Learning Sciences, pp. 880–886.

Zuckerman, O., Resnick, M., 2003. System blocks: a physical interface for system dynamics simulation. In: Proceedings of CHI'03, pp. 810–811.

Zuckerman, O., Resnick, M., 2004. Hands-on modeling and simulation of systems. In: Proceedings of IDC'04, pp. 157–158.

Zuckerman, O., Resnick, M., 2005. Children's misconceptions as barriers to learning stock- and flow- modeling. In: Proceedings of the 23rd International Conference of the System Dynamics Society, 2005.