A Tiered Evaluation Framework for Reality-Based Creativity Support Environments

Orit Shaer, Consuelo Valdes
Wellesley College
106 Central St, Wellesley, MA
{oshaer, cvaldes} @wellesley.edu

ABSTRACT
Reality-based interfaces (RBIs) such as tabletop and tangible user interfaces offer unique opportunities for supporting creativity. In this paper, we present a multi-tiered evaluation framework for reality-based creativity support environments and describe its application in the context of college-level science education. The proposed framework consists of three layers that examine the usability, usefulness, and impact of an environment on creative problem solving processes. Drawing upon the existing body of work in the area, our framework documents a mixed-method approach and provides guidance for the evaluation of reality-based creativity support environments.

Author Keywords
Reality-based interaction; creativity support; tabletop interaction; collaborative learning.

ACM Classification Keywords
H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous.

General Terms
Human Factors; Design; Measurement.

INTRODUCTION
Over the past two decades, HCI research has generated a broad range of interaction styles that move beyond the desktop into new physical and social contexts. Key areas of innovation have been tabletop, tangible, and embodied user interfaces. These interaction styles share an important commonality: leveraging users' existing knowledge and skills of interaction with the real non-digital world, thus they can be unified under the umbrella of Reality-Based Interfaces (RBIs) [4]. Building upon ideas from embodied cognition, RBIs offer a more natural, intuitive, and accessible form of interaction that reduces the mental effort required to learn and operate a computational system [4].

Given the potential of RBIs, numerous research prototypes have explored how these emerging interaction styles will impact education. Several studies have examined the effects of RBIs on learning, investigating the benefits and deficits of reality-based interaction in the contexts of formal and informal learning. However, most of these studies have focused on children. To date, little research has been devoted to investigating the strengths and limitations of utilizing RBIs for promoting creative thinking in scientific problem solving at the high school or college levels.

Our focus is on investigating the application of reality-based interaction for promoting creative thinking in college-level science education. Creative thinking in the context of science education is a multicomponent process, which is influenced by group interactions and social context [1]. The creative process that leads to an individual insight (i.e. unit of discovery [9]) includes at least three diverse, but testable elements [1]: 1) divergent thinking – the ability to generate and accept many ideas related to a problem; 2) convergent thinking – the ability to focus and mentally evaluate ideas; and 3) analogical thinking – the ability to understand a novel idea in terms of existing knowledge. In our work, we have explored the application of reality-based interaction for enhancing creative thinking in data-intensive areas such as college-level genomics [11, 12, 13] and phenology [15].

In this position paper we describe a multi-tiered evaluation framework for understanding the strengths and limitations of reality-based creativity support environments [13]. We applied this framework in the evaluation of three reality-based interfaces for inquiry-based college level science learning: G-nome Surfer [11, 12, 13] – a tabletop user interface for collaborative exploration of genomic information; GreenTouch [15] – a collaborative environment for engaging novices in scientific inquiry in phenology; and MoClo Planner – a multi-touch interface for collaborative bio-design. Our evaluation framework consists of three layers, which examine the *usability*, *usefulness*, and *impact* of reality-based interaction on creative problem solving in a collaborative context. Drawing upon the existing body of work in the area, our framework documents a mixed-method approach that aims to provide guidance (rather than an extensive checklist) for the evaluation of reality-based creativity support environments.

CHI'13, April 27 – May 2, 2013, Paris, France.

Following, we describe our evaluation framework. We then demonstrate its application in the evaluation of G-nome Surfer 2.0.

MULTI-TIERED EVALUATION FRAMEWORK

The proposed framework consists of three layers that examine the *usability*, *usefulness*, and *impact* of reality-based creativity support environments. The first layer applies a micro perspective – focusing on the usability of concrete interaction techniques and the effectiveness of individual visualizations. The second layer applies a macro perspective – studying the usefulness of a system in the context of a full-scale task. Finally, the third layer applies a holistic perspective – examining the impact of the system on users' performance in-situ. Table 1 provides a summary of our evaluation framework. For each layer, we describe its dimensions, settings, metrics, and methods for data collections.

		L1: Usability	L2: Usefulness	L3: Impact
	Dimensions	Functionality Learnability Performance Memorability Errors Satisfaction	Effectiveness Efficiency	Performance Engagement Collaboration
	Settings	In Lab Study	Comparative Study	In-Situ deployment Longitudinal evaluation
	Metrics			
Task	Completion	x	x	x
	Workload	x	x	x
	Accuracy	x	x	x
	Time on task	x		
	Number and type of errors	x		
User	Attitudes	x	x	
	Satisfaction ratings	x		
	Levels of participation		x	
	Equity of participation		x	
	Engagement ratings			x
Learning	Nature of discussion		x	x
	Nature of collaboration		x	x
	Problem solving strategies			x
	Number of hypotheses			x
	Methods			
	Expert review	x		
	Logging information	x		
	Observation	x	x	x
	Discourse analysis		x	x
	Video coding		x	x
	Questionnaires	x	x	x
	Interviews		x	x
	Debrief	x		

Table 1: A multi-tier evaluation framework for collaborative tabletop interaction.

The first layer, *usability* (L1), consists of six dimensions that draw upon Schneiderman's definition of usability [14] and are not specific for reality-based interaction. These include: 1) functionality – the ability of the system to support the user in completing a required set of tasks; 2) learnability – the extent to which it is easy to learn how to use a system; 3) performance – the extent to which the accomplishment of a task satisfies known standards of completeness, accuracy, and speed; 4) memorability – the ability of the user to re-establish proficiency using a system after a period of inactive use; 5) errors – the frequency, type, and severity of errors as well as how easy it is to recover from errors; and 6) satisfaction – the degree to which a user finds the system pleasant to use. These dimensions are easily quantifiable using mostly task-centered metrics as specified in Table 1.

The second layer examines *usefulness* (L2), the advantages of a system for accomplishing creative tasks in a collaborative setting. It consists of two dimensions: 1) effectiveness – the extent to which users' goals are obtained through an effective collaborative process, a process where group members actively communicate with each other to demonstrate shared effort [5]; and 2) efficiency – the degree to which goals are obtained with the investment of less time as well as physical and mental effort. Effectiveness and efficiency are interdependent and should be considered together.

These dimensions can be quantified by combining various task-, user-, and learning-centered metrics that are calculated using mixed methods. For example, the effectiveness of a collaborative tabletop interface can be quantified by measuring task completion rates along with rating the effectiveness of the collaborative process. Collaboration profiles [12] are often useful for describing the nature of collaboration by highlighting the different roles participants assume throughout the collaborative process. Computing the level of participation per user is helpful for calculating the equity of participation [2]. Dialog analysis can provide further insight into the nature of discussion carried by users while working on a task – helping to identify divergent and convergent thinking. For example, such dialog analysis can reveal the time spent on task-related vs. non task-related talk while highlighting insights gained by users [9] and reflective utterances (which indicate analogical thinking). Efficiency can be quantified by measuring task completion time as well as mental and physical effort. Subjective mental and physical effort is often measured using the standard NASA TLX questionnaire [3].

Finally, the third layer focuses on studying the *impact* (L3) of a creativity support system on users' performance and practices in-situ. This layer takes a holistic approach, studying impact on three dimensions: 1) performance – here we consider performance more broadly than in the usability layer, examining not only quantitative task-centered metrics such as time, completion rates, and workload, but also creativity-centered measures that focus on how users apply creative thinking. In particular, we look into how users solve problems in collaborative settings. We suggest utilizing video and discourse analysis to identify behavioral profiles, problem solving strategies, and the number and quality of hypotheses (i.e. alternative solutions) explored by users; 2) engagement – this dimension goes beyond mere user satisfaction to capture the degree of user's interest, emotional involvement, and dynamic interaction. O'Brien et al. (2008) developed a multi-scale measure for user engagement that considers six attributes of engagement: Perceived Usability, Aesthetics, Focused Attention, Felt Involvement, Novelty, and Endurability; and 3) collaboration – the degree and manner to which users collaborate on a task. Here, we consider various metrics that indicate how users collaborate, what roles they assume during the collaborative process, and whether and when they switch their roles. To rate the effectiveness of the collaborative process, we use the rating scheme created by Meier's' et al [7] that considers 5 different dimensions of collaboration: communication, information pulling, coordination, interpersonal relationship, and motivation.

Taken together these three layers consider creative thinking as a multicomponent process that is strongly influenced by group interactions [1].

Following, we describe briefly the application of this framework to the evaluation of G-nome Surfer. A detailed report describing the evaluation process, results, and findings is provided in [11, 12, 13].

EVALUATING G-NOME SURFER

G-nome Surfer [11, 12, 13] is a tabletop user interface for collaborative exploration of genomic information (see Figure 1). It was designed to support hypotheses forming by facilitating collaborative, immediate, and fluid interaction with large amounts of heterogeneous genomic information. G-nome Surfer utilizes multi-touch and tangible interaction techniques to lower the threshold for using advanced bioinformatics tools as well as to provide support for the divergent, convergent, and analogical stages of the creative inquiry process.

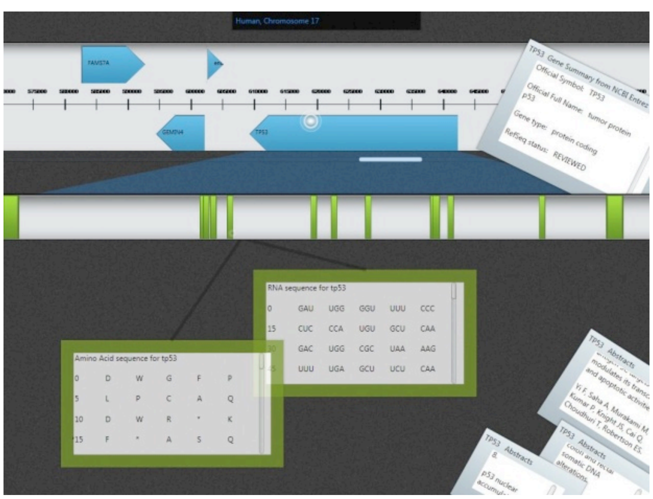

Figure 1: G-nome Surfer 1.0 displaying the human gene TP53 and related publications.

We applied the first layer of evaluation, *usability* (L1), on a continual basis throughout the development of G-nome Surfer. In addition to user testing of each of the complete versions, we often conducted micro-studies examining the usability of particular features through the iterative development and testing of a series of prototypes in increasing fidelity [11].

The second layer, the evaluation of usefulness (L2), was applied through an experimental study with 48 participants that compared undergraduate students' learning of genomics using existing bioinformatics tools (i.e. GUI condition) and two alternative prototypes of G-nome Surfer 2.0: a collaborative multi-mouse GUI and a tabletop interface [12]. We also evaluated the usefulness of G-nome Surfer Pro through a study with 14 student researchers that used the interface for microbiology research.

Our findings highlight several advantages of tabletop interaction for creative problem solving compared to multi-mouse GUI including: 1) *reflective dialogue* (i.e. analogical thinking) – in the tabletop condition, participants spent significantly more time on reflective activities and articulated a larger number of insights than in the other conditions; 2) *physical participation* – participants in the tabletop condition exhibited significantly higher levels of physical participation, expressed by increased spatial manipulation of information. We observed that in the tabletop condition participants manipulated information artifacts – moving, resizing, and rotating – to a greater extent than in the two other conditions. Often, users aligned information artifacts side by side for comparison and then moved them around the table to share with their partner or to place them in an area of the tabletop for later use; 3) *intuitive interaction* – the tabletop condition facilitated more intuitive interaction. This was evident from a statistically significant lower number of utterances related to interaction syntax compared to the two other conditions, and from the reduced time spent on finding information rather than discussing it; and 4) *effective collaboration* – in the tabletop condition, participants were engaged in a more effective collaboration than in the other conditions. This was evident from the turn-taking collaboration style exhibited by most tabletop pairs. Discourse analysis data revealed that in the tabletop condition there were a significantly higher number of coordination utterances and a significantly lower number of disengagement utterances compared to the multi-mouse GUI. Taken together, turn-taking style, higher number of coordination utterances, and lower number of disengagement utterances indicate effective collaboration.

Finally, we applied the third layer of our evaluation framework to study the *impact* (L3) of G-nome Surfer 2.0 in authentic educational settings, deploying it in an intermediate-level undergraduate Neuroscience laboratory course at our institution. Results from this evaluation provide empirical evidence for the feasibility and value of integrating reality-based creativity support environments in college-level science education as well as shed light on how users collaborate and solve problems using such environments in the context of college level inquiry-based learning [13].

Figure 2: G-nome Surfer comparing gene ontology and expression data of different mouse genes.

CONCLUSION

We have presented a multi-tiered evaluation framework for reality-based creativity support environments and described its application in the context of college-level inquiry based learning. The proposed framework consists of three layers that examine the *usability*, *usefulness*, and *impact* of an environment on creative problem solving. This framework takes a holistic approach for gaining an understanding of the strengths and limitations of an environment by utilizing a variety of quantitative measures and qualitative indicators. Combined together, the dimensions and metrics proposed by this framework highlight multiple facets of the creative process mediated by a particular interface.

ACKNOWLEDGMENTS

This work is partially funded by NSF grant IIS-1017693.

REFERENCES

1. Dehaan, R. L. (2009). Teaching Creativity and Inventive Problem Solving in Science. *CBE Life Science Education*, 8(3):172-181.
2. Harris, A., Rick, J., Bonnett, V., Yuill, N., Fleck, R., Marshall, P., et al. (2009). Around the table: Are multiple-touch surfaces better than single-touch for children's collaborative interactions? *ACM Computer-Supported Collaborative Learning* (pp. 335-344). ACM Press.
3. Hart, S., & Staveland, L. (1988). Development of NASA-TLX (Task Load Index) : Results of Empirical and Theoretical Research. *Human Mental Workload* .
4. Jacob, R. J., Girouard, A., Hirshfield, L., Horn, M. S., Shaer, O., Solovey, E. T., et al. (2008). Reality-Based Interaction: A Framework for Post-WIMP Interfaces. *Human Factors in Computing Systems* . Florence: ACM.
5. Kirschner, F., Pass, F., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains from complex learning. *Educational Psychology Review*, 31-42.
6. Liu, S., Lu, K., Seifeselassie, N., Grote, C., Francisco, N., Lin, V., et al. (2012). MoClo Planner: Supporting Innovation in Bio-Design through Multitouch Interaction. *Interactive Tabletops & Surfaces.* Cambridge, MA: ACM.
7. Meier, A., Spada, H., & Rummel, N. (2007). A rating system for assesing the quality of computer-supported collaboration processes. *International Journal of Computer-Supported Collaborative Learning*, 63-86.
8. O'Brien, H. L., & Toms, E. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science and Technology*, 938-955.
9. Saraiya, P., North, C., & Duca, K. (2005). An Insight-based methodology for evaluating bioinformatics visualizations. *Visualization and Computer Graphics*, 443-456.
10. Schneider, B., Strait, M., Muller, L., Elfenbein, S., Shaer, O., & Shen, C. (2012). Phylo-Genie: Engaging Students in Collaborative 'TreeThinking' through Tabletop Techniques. *Human Factors in Computing Systems* . Austin: ACM.
11. Shaer, O., Kol, G., Strait, M., Fan, C., Catherine, G., & Elfenbein, S. (2010). G-nome Surfer: a Tabletop Interface for Collaborative. *Conference on Human Factors in Computing Systems.* 2010: ACM.
12. Shaer, O., Strait, M., Valdes, C., Feng, T., Lintz, M., & Wang, H. (2011). Enhancing Genomic Learning through Tabletop Interaction. *Human Factors in Computing Systems* . Vancouver: ACM.
13. Shaer, O., Strait, M., Valdes, C., Wang, H., Feng, T., Lintz, M., et al. (2012). The design, development, and deployment of a tabletop interface for collaborative exploration of genomic data. *International Journal of Human-Computer Studies*, 746-764.
14. Shneiderman, B., & C., P. (2008). *Designing the User Interface, 5th Ed.* Addison-Wesley.
15. Valdes, C., Ferreirae, M., Feng, T., Wang, H., Tempel, K., Liu, S., et al. (2012). A Collaborative Environment for Engaging Novices in Scientific Inquiry. *Interactice Tabletops and Surfaces.* Cambridge: ACM