The Impact of Interface Affordances on Human Ideation, Problem-solving and Inferential Reasoning

SHARON OVIATT, Incia Designs
ADRIENNE COHEN, Duke University
ANDREA MILLER, Stanford
KUMI HODGE, Stanford
ARIANA MANN, MIT

Two studies investigated how computer interface affordances influence basic cognition, including ideational fluency, problem solving, and inferential reasoning. In one study comparing interfaces with different input capabilities, students expressed 56% more nonlinguistic representations (diagrams, symbols, numbers) when using pen interfaces. A linear regression confirmed that nonlinguistic communication directly mediated a substantial increase (38.5%) in students' ability to produce appropriate science ideas. In contrast, students expressed 41% more linguistic content when using a keyboard-based interface, which mediated a drop in science ideation. A follow-up study pursued the question of how interfaces that prime nonlinguistic communication so effectively facilitate cognition. This study examined the relation between students' expression of nonlinguistic representations and their inference accuracy when using analogous digital and non-digital pen tools. Perhaps surprisingly, the digital pen interface stimulated construction of more diagrams, more correct Venn diagrams, and more accurate domain inferences. Students' construction of multiple diagrams to represent a problem also directly suppressed overgeneralization errors, the most common inference failure. These research results reveal that computer interfaces have communications affordances, which elicit communication patterns that can substantially stimulate or impede basic cognition. Implications are discussed for designing new digital tools for thinking, with an emphasis on nonlinguistic and especially spatial representations that are most poorly supported by current keyboard-based interfaces.

Categories and Subject Descriptors: H5.2 Information interfaces and presentation: User interfaces- User-centered design, input devices and strategies, evaluation/methodology
General Terms: Experimentation, Design, Performance, Human Factors
Additional Key Words and Phrases: Pen interfaces, Educational interfaces, Thinking tools, Ideational fluency, Problem solving, Inferential reasoning, Nonlinguistic representations, Diagrams, Affordances

ACM Reference Format:
DOI =

1. INTRODUCTION

A central theme of this article is that expressively rich interfaces have the potential to stimulate ideation, problem solving, and inferential reasoning. In particular, input capabilities that encourage expressing information in different representations, especially nonlinguistic ones (e.g., numbers, symbols, diagrams), can facilitate clarity of thought. This is especially important in designing effective tools for education, since educational interfaces need to facilitate students’ own engagement, communication, and problem-solving activities, including increasing their communicative activity in a way that directly stimulates thought.

1.1 Limitations of Current Keyboard-and-Mouse Graphical Interfaces

Existing keyboard-based graphical interfaces limit users' ability to fluidly express different representations, especially diagrams and symbols. In educational contexts, this is a handicap, because it limits the flow of thinking through steps during problem-solving activities [Oviatt, Arthur, Brock and Cohen 2007]. As a result, although existing keyboard-and-mouse graphical interfaces are well suited for handling relatively mechanical tasks, such as email, text editing, and web search, they can be less effective for supporting extended human thought and learning [Crowne 2007; Oviatt, Arthur and Cohen 2006; Oviatt and Cohen 2010a]. Unlike
newer input alternatives such as digital pens, keyboards are particularly unsuitable for expressing spatial information, even though cognitive science research indicates that human cognitive abilities have a spatial foundation:

“The evidence indeed suggests that human reasoners use functionally spatial models to think about space, but they also appear to use such models in order to think in general.” —Johnson-Laird [1999, p. 60]

When students use a keyboard-based interface, recent research indicates that they experience higher cognitive load that can undermine their performance during STEM tasks (science, technology, engineering, mathematics). During science activities, students who used a keyboard-based graphical interface solved problems more slowly, solved fewer problems correctly, and forgot more problem content, compared with using a digital pen interface [Oviatt and Cohen 2010a]. For the same students solving the same science problems, their percentage of correct solutions dropped 10% when using a keyboard-based interface, or a whole grade point.

During mathematics tasks, a comparison of students’ speed, attention, metacognitive control, correctness of solutions, and memory revealed that they performed better when using a digital pen and paper interface than a pen tablet interface, which in turn supported better performance than a keyboard-based graphical interface [Oviatt, Arthur et al. 2006]. In particular, think-aloud protocols revealed that students’ ability to focus attention deteriorated substantially when using the keyboard interface, which undermined their ability to think in a high-level strategic manner about how to solve problems. When using this interface, a 50% drop occurred in think-aloud comments about what type of math problem they were working on (e.g., “Oh, it’s a 3-D problem, not a 2-D one”), strategies needed to solve the problem, its difficulty level, whether they were in an error state, and other information required to self-regulate performance successfully [Oviatt, Arthur, et al. 2006]. Some metrics revealed that lower-performing students were more adversely affected by the keyboard interface than high performers. For example, low performers’ ability to solve math problems correctly and to remember the problem content declined more substantially than that of high performers when using the keyboard interface. In summary, recent studies involving converging metrics all have documented performance deterioration when students use keyboard-based interfaces during STEM tasks, compared with non-digital tools or pen interfaces.

In contrast, pen interfaces provide a single digital tool for expressing and shifting among all the different representations (i.e., symbols, digits, diagrams, words) during the flow of problem-solving steps. For example, if a geometry problem is presented as a word problem, a student could first diagram the relation between objects, then generate algebraic expressions to solve the problem using symbols and numbers, and end by summarizing their answer using linguistic content. This content could be expressed fluidly, without interrupting thought, using a single pen interface. Improved flow would reduce working memory demands, thereby lowering cognitive load. The properties described above mean that pen interfaces have greater expressive range, precision, and flexibility than a keyboard-based interface. If communication mediates thought, these characteristics have the potential to stimulate increased mental effort and the refinement of ideas needed to achieve correct solutions.

1.2 Hypothesis Generation, Ideational Fluency and Scientific Innovation

The ability to ask good questions and to generate related hypotheses is an important hallmark of scientific thought. Hypothesis generation has been underacknowledged as a critical precondition for successful problem solving, including decision-making in
real-world tasks such as medical and mechanical diagnosis. For example, if a doctor or mechanic cannot generate the correct cause for an observed phenomenon among the hypotheses being considered, then decision-making will fail. Unfortunately, decision makers typically only generate a small set of high-likelihood hypotheses, are overconfident in their completeness, and unlikely to consider additional hypotheses beyond their initial set [Barrows, Norman, Neufeld and Feightner 1982; Ruths, Nakhlhe, Iyengar, Reddy and Ram 2006; Weber, Bockenholt et al. 1993].

The ability to produce many appropriate hypotheses within a domain demonstrates ideational fluency, a concept based on Guilford and colleagues' research on divergent thinking and creativity [Guilford 1956]. In this study, it was scored as the total number of distinct and appropriate biology hypotheses that students produced (e.g., “Toes on chimp foot adapted for grasping, like human hand” would count as one hypothesis in response to problem 2 in Figure 2).

### Table 1. Definition of Terms & Dependent Measures

<table>
<thead>
<tr>
<th>Term</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Linguistic fluency</strong></td>
<td>is the total number of words, abbreviations, and acronyms that students produced while working on hypothesis-generation or problem-solving tasks.</td>
</tr>
<tr>
<td><strong>Nonlinguistic fluency</strong></td>
<td>is the total number of digits, symbols (e.g., →), diagrams (e.g., Punnett square), and thinking marks (see below) that students produced while working on hypothesis-generation or problem-solving tasks.</td>
</tr>
<tr>
<td><strong>Communicative fluency</strong></td>
<td>is the total linguistic and nonlinguistic fluency expressed involving all representational systems, as used previously in the literature (Oviatt, Arthur et al., 2007).</td>
</tr>
<tr>
<td><strong>Clausal constructions</strong></td>
<td>depart from full sentences by omitting a subject or verb constituent (e.g., “Development of a justice system” as a student response to problem 1 in Figure 2).</td>
</tr>
<tr>
<td><strong>Thinking marks</strong></td>
<td>are the total number of pen marks that students made on visuals displayed in problem statements as an aid in counting, selecting, ordering, grouping, labeling, or showing relations between information elements. Students make these marks as they clarify a problem’s meaning and self-organize information elements in preparation for solving it (Oviatt &amp; Cohen, 2010b). Marking typically precedes diagramming during problem solving.</td>
</tr>
<tr>
<td><strong>Schemas</strong></td>
<td>are mental constructions or representations of learned information that structure our understanding of the world.</td>
</tr>
<tr>
<td><strong>Ideational fluency</strong></td>
<td>is a metric based on Guilford and colleagues' research on intellectual activity during divergent thinking (Guilford, 1956). In this study, it was scored as the total number of distinct and appropriate biology hypotheses that students produced (e.g., “Toes on chimp foot adapted for grasping, like human hand” would count as one hypothesis in response to problem 2 in Figure 2).</td>
</tr>
</tbody>
</table>

Affordance theory maintains that people have perceptually-based expectations about objects, such as computer interfaces, including beliefs about constraints on how you can interact with them successfully. These affordances establish behavioral attunements that transparently but powerfully prime the likelihood that people will act on objects in specific ways that differentiate them [Gibson 1977]. Importantly, the behavioral attunements that arise from interface affordances are distinct from specific learned patterns. Affordances potentially can elicit exploratory activity that facilitates new learning. As a result, if interface affordances are well matched with a task domain, they can increase physical or communicative activity patterns that facilitate the construction of new schemas. See Table 1 for terms. They also can act as catalysts for transferring procedural and related domain knowledge across different tasks and contexts.

The present research extends Affordance theory to predict that people also will engage in specific patterns of **communicative activity** when they use interface tools with different input capabilities, such as a keyboard or pen. It explores whether the presence of a computer interface will elicit more total communicative activity from people than a non-digital pencil and paper tool. It also investigates whether interfaces
characterized by different input capabilities have *communications affordances* that prime qualitatively different types of communication pattern. In particular, it was predicted that pen interfaces would selectively stimulate increased nonlinguistic communicative fluency, such as informal marking, diagrams, numbers, and symbols. In contrast, it was expected that keyboard-based graphical interfaces would selectively prime increased linguistic fluency, and also full sentential constructions for presentation purposes. See Table 1 for terms.

An additional objective was to examine whether an increase in communicative activity would likewise stimulate mental effort, the production of ideas, and the correctness of problem solving outcomes [Sweller, van Merrienboer and Paas 1998]. This type of facilitation would depend on whether the representations used during communication are well matched with a task domain. As background, Activity theory maintains that language plays a major role in mediating, guiding, and refining mental activities, including planning and problem solving [Luria 1961; Vygotsky 1962]. For example, people spontaneously engage in self-talk as they work on difficult tasks, and it is documented to be an effective strategy for guiding improved performance [Luria 1961; Vygotsky 1962]. We likewise expect that active marking in the written modality can function to guide improved performance. Activity theory therefore provides a basis for predicting that interface tools that prime communicative activity can stimulate a parallel increase in ideational fluency and also serve to refine the accuracy of ideas.

2. STUDY ONE: THE IMPACT OF INTERFACE AFFORDANCES ON IDEA GENERATION AND PROBLEM SOLVING

In the present research, it was hypothesized that when students actively generate more nonlinguistic content while solving science problems, their problem solution scores would improve. Nonlinguistic content includes symbols and numbers, but also spatial content such as diagrams and “thinking marks” that students make to organize information displayed in problem statements (see terms in Table 1). It also was predicted that pen interfaces, which support expressing different nonlinguistic representations [Oviatt, Arthur et al. 2007], would stimulate increased ideational fluency in the form of scientific hypothesis generation. In contrast, keyboard interfaces that prime linguistic fluency, but have limited support for nonlinguistic representations, would be associated with decreased ideational fluency during hypothesis generation tasks.

In summary, it was predicted that the (1) presence or absence of a computer interface, and (2) type of computer interface used would systematically influence both the quantity and content of students’ communicative fluency. In addition, (3) if an interface is well matched with a task domain, then it will facilitate students’ ability to generate appropriate ideas and solve problems correctly. The compatibility of an interface with a task domain depends on whether it supports expression of representations that are required to solve domain problems (e.g., diagramming of Punnett squares for genetics inheritance problems). These three main study predictions are associated with 22 specific hypotheses and planned contrasts, which are outlined in Table 2.

2.1 Study One: Methods

Sixteen high-school biology students participated as paid volunteers in a longitudinal study comparing interfaces for science education. All had recently completed an introductory biology class and expressed an interest in technology. To ensure that new interfaces are designed for diverse students, eight participants were high performing and eight low performing according to year-end biology grades, with each subgroup gender balanced. All participants were native speakers of English, and used non-
C. Keyboard Interfaces Stimulate More Linguistic Communicative Fluency

A-H1: Compared with paper and pencil (PP), digital interfaces (DP, PT, GT) will stimulate higher total communicative fluency (linguistic and nonlinguistic) during hypothesis-generation tasks
A-H2: Compared with paper and pencil (PP), digital interfaces (DP, PT, GT) will stimulate higher total communicative fluency during problem-solving tasks
A-H3: Compared with paper and pencil (PP), pen interfaces (DP, PT) will prime higher nonlinguistic communicative fluency during hypothesis-generation tasks (Figure 4A)
A-H4: Compared with paper and pencil (PP), pen interfaces (DP, PT) will prime higher nonlinguistic communicative fluency during problem-solving tasks
A-H5: Compared with paper and pencil (PP), a keyboard-based graphical interface (GT) will prime higher linguistic communicative fluency during hypothesis-generation tasks (Figure 4B)
A-H6: Compared with paper and pencil (PP), a keyboard-based graphical interface (GT) will prime higher linguistic communicative fluency during problem-solving tasks

B. Pen Interfaces Stimulate More Nonlinguistic Communicative Fluency

B-H7: Compared with paper and pencil or a keyboard interface (PP, GT), pen interfaces (DP, PT) will prime higher nonlinguistic communication during hypothesis-generation tasks
B-H8: Compared with paper and pencil or a keyboard interface (PP, GT), pen interfaces (DP, PT) will prime higher nonlinguistic communication during convergent problem-solving tasks
B-H9: Compared with paper and pencil or a keyboard interface (PP, GT), pen interfaces (DP, PT) will prime higher nonlinguistic communicative fluency indexed as thinking marks during convergent problem-solving tasks
B-H10: Compared with paper and pencil or a keyboard interface (PP, GT), pen interfaces (DP, PT) will prime higher nonlinguistic communication for low-performing students during hypothesis-generation and problem-solving tasks
B-H11: Compared with paper and pencil or a keyboard interface (PP, GT), pen interfaces (DP, PT) will prime higher nonlinguistic communication for high-performing students during hypothesis-generation and problem-solving tasks
(Note: Also see A-H3 and H4)

C. Keyboard Interfaces Stimulate More Linguistic Communicative Fluency

C-H12: Compared with non-keyboard interfaces (PP, DP, PT), a keyboard-based interface (GT) will prime higher linguistic communicative fluency during hypothesis-generation tasks
C-H13: Compared with non-keyboard interfaces (PP, DP, PT), a keyboard-based interface (GT) will prime higher linguistic communication during problem-solving tasks
C-H14: Compared with non-keyboard interfaces (PP, DP, PT), a keyboard-based interface (GT) will prime higher linguistic communicative fluency for both low-performing students during hypothesis-generation and problem-solving tasks
C-H15: Compared with non-keyboard interfaces (PP, DP, PT), a keyboard-based interface (GT) will prime higher linguistic communicative fluency for both high-performing students during hypothesis-generation and problem-solving tasks
C-H16: Compared with non-keyboard interfaces (PP, DP, PT), a keyboard-based interface (GT) will prime a higher ratio of full sentences to clausal constructions during hypothesis-generation tasks (Figure 7)
(Note: Also see A-H5 and H6)

D. Nonlinguistic Communicative Fluency Facilitates Scientific Idea Generation & Improved Problem Solving

D-H17: Compared with paper and pencil (PP), pen interfaces (DP, PT) will stimulate higher ideational fluency indexed as science hypothesis generation (Figure 5A)
D-H18: Compared with paper and pencil or a keyboard interface (PP, GT), pen interfaces (DP, PT) will stimulate higher ideational fluency indexed as science hypothesis generation
D-H19: During hypothesis-generation, interfaces that effectively prime nonlinguistic communicative fluency will directly facilitate higher ideational fluency (Figure 5B)
D-H20: Active marking in diagrams will be associated with higher scores on science problem solving than not diagramming, given matching on problem, student, interface used (Figure 6A)
D-H21: Active marking in thinking marks will be associated with higher scores on science problem solving than not marking, given matching on problem, student, interface used (Figure 6B)

E. Linguistic Communicative Fluency Suppresses Scientific Idea Generation

E-H22: During science hypothesis generation, interfaces that elicit greater linguistic communicative fluency will strongly suppress ideational fluency (Figure 8)
Fig. 1. Toshiba laptop screen with biology problem display, and keyboard and stylus for input (left, bottom); Student response that combines pen input and typing, using the GT interface and OneNote (left, top); Student response with pen input, using DP interface and Maxell pen (right).

digital tools in their biology classes. However, they all were experienced users of keyboard-and-mouse graphical interfaces for other general purposes. In contrast, only two students had ever used a pen interface before, and none were experienced regular users.

Students were asked to complete twelve idea-generation problems using: (1) non-digital paper and pencil (PP), (2) a digital pen and paper interface (DP) (Anoto, 2010), (3) a pen tablet interface (PT), and (4) a graphical tablet interface (GT) with keyboard, mouse and pen. Figure 1 illustrates the interface materials. In all conditions, problems were read on a laptop screen as shown in Figure 1 (lower left). Figure 1 (lower left). In the two paper conditions, students entered their work on paper, while in the two tablet conditions they worked on the computer using OneNote note-taking software [OneNote 2011]. In all conditions, students had comparable white space. When problems included a diagram or table, these visuals were available in students’ workspace so they could write or type directly on them while working.

Students also had access to a writing implement for input in all conditions, as summarized in Table 3: a pencil (PP), digital pen (DP), tablet stylus (PT), and free choice to use stylus, keyboard, and mouse in any way they wished (GT). From an objective standpoint, the hardware provided in all conditions provided equal support for drawing diagrams or symbols easily. In the case of the free choice graphical interface (GT), one goal of this study was to evaluate whether students’ communication patterns when using it would be dominated by the keyboard. Although interface designers may assume that people will use different input capabilities in an optimal way to support their own performance, this may not be the case. If users
perceive that the dominant GT interface affordance is keyboard-centric, then they may use keyboard input even if it does not support their performance best.

Table 3. Summary of interfaces and their primary input, output and workspace features

<table>
<thead>
<tr>
<th>Interface Type</th>
<th>Problem Statement†</th>
<th>Input Tools</th>
<th>Work Space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pencil &amp; Paper (PP)</td>
<td>Tablet screen</td>
<td>Pencil</td>
<td>Hardcopy Paper</td>
</tr>
<tr>
<td>Pen Tablet (PT)</td>
<td>Tablet screen</td>
<td>Tablet stylus</td>
<td>OneNote on Computer Screen</td>
</tr>
<tr>
<td>Graphical Tablet (GT)</td>
<td>Tablet screen</td>
<td>Keyboard, mouse &amp; Tablet stylus†</td>
<td>OneNote on Computer Screen</td>
</tr>
</tbody>
</table>

† All problems were presented visually on a Toshiba or Fujitsu laptop screen, with students pressing a submit button after completing each problem to view the next one.
‡ Students had free choice to use the keyboard, mouse or pen as they wished.

Fig. 2. Examples of hypothesis-generation tasks

Students worked on hypothesis-generation problems, such as (1) figures summarizing biological trends, for which they were instructed to provide as many hypotheses as possible to explain them, and (2) comparative biological structures, for which they were asked to generate possible reasons why they differed. Examples are listed in Figure 2. Problem sets were composed based on consultation with biology teachers and also pilot testing. The problems involved diverse content to ensure generalizability of any results. Each student completed three hypothesis-generation tasks in each of the four interface conditions, or twelve tasks total.

Students worked individually at their own pace for approximately one hour while producing ideas. Data on these hypothesis generation tasks were collected during the second session in a longitudinal sequence, so participants already had received orientation, practice, and 1-2 hours of experience using each interface. During this second session, they were given instructions and practice on how to use the basic
interface features required to complete their problems. For example, with the tablet interfaces they were shown how to ink, erase, undo/redo, move and resize input, and scroll down to get more writing space. With the digital pen interface, they were shown the Maxell pen’s vibro-tactile feedback when: (1) removing the cap to start the pen computer, (2) making a checkmark in the “Done” box (Figure 1, lower right) to transmit a completed answer to the laptop, and (3) processing problems occurred, such as low battery. They also were shown how to ink, erase, and redo any input while working. Before starting, students were instructed to take their time and concentrate on producing as many appropriate hypotheses as they could for each problem.

Within-subject data were collected on the same students’ idea generation for the same problems in order to provide a sensitive assessment of differences associated specifically with using alternative interfaces. The serial position of each interface during within-subject testing, and its pairing with different problem sets, were considered nuisance variables. They were completely counterbalanced within each group of four students. Analyses of the impact of different interfaces on the main dependent measures were conducted on the fully counterbalanced subject sets. A priori planned analyses focused on measures of communicative fluency and ideational fluency associated with specific study hypotheses. See Table 1 for definition of terms and dependent measures, and Table 2 for summary of hypotheses and related planned contrasts.

Additional analyses compared the impact of diagramming and informal marking on students’ solution correctness, based on matched pairs of convergent problem-solving tasks for which students did or did not make a diagram or informal marks. A sample problem-solving task is shown in Figure 3. Typical problem content and other methodological details are summarized in Oviatt and Cohen [2010a].

During the study, 192 hypothesis-generation and 512 problem-solving tasks were available for analysis, or 704 problems total. The total linguistic and nonlinguistic communication content available from these tasks was approximately 6,700 and 1,250 items, respectively. As validation of performance differences between the student groups, high-performing students averaged 78.1% on biology problem-solving tasks, whereas low-performing students averaged 57.5%.

Analyses confirmed that 65% of all student input when using the GT interface was communicated using the keyboard, even though they had free choice to use the keyboard or stylus whenever they wished. These data validate that the GT interface was “keyboard-centric,” or the dominant input mode with respect to interface affordances influencing students’ activity pattern.

All communicative fluency measures (see Table 1) were second scored by an independent rater, and averaged 95.5% reliability across different types of representation (range 80.4%-100%). All hypothesis-generation and problem-solving responses were scored in full by two independent expert raters, with any departures resolved by a third scorer who was an expert biology teacher.

![Fig. 3. Genetic inheritance problem](image-url)
2.2 Study One: Results

2.2.1. Computer Interfaces Stimulate More Communicative Fluency. As predicted, when students used the computer interfaces (DP, PT, GT) to generate biology hypotheses, their total communicative fluency was 15.5% higher than when they used non-digital paper and pencil tools (PP) (\(x\̄= 43.6\) v. 37.8, respectively), paired \(t = 2.87\) (df 3) \(p < 0.032\), one-tailed. This finding replicated when they were engaged in problem-solving tasks, with a 23.5% higher total communicative fluency (\(x\̄= 36.8\) v. 29.8, respectively), paired \(t = 2.78\) (df 3) \(p < 0.034\), one-tailed.

Four additional analyses examined more specific initial study predictions that: (1) pen interfaces (DP, PT) would elevate nonlinguistic communicative fluency more than non-digital pencil and paper tools (PP), and (2) the keyboard-based interface (GT) would elevate linguistic fluency more than non-digital tools (PP). Analyses were conducted separately for nonlinguistic and linguistic fluency mainly to evaluate these differential predictions. They also were evaluated separately because these two types of communicative fluency reflect different scales. Whereas the number of words expressed per problem averaged 30-45, the number of nonlinguistic content items (e.g., diagrams) averaged 5-8.

As shown in Figure 4A, these more specific analyses confirmed the prediction that pen interfaces increase students’ nonlinguistic communication more than non-digital pencil and paper tools during hypothesis generation (\(x\̄= 7.46\) v. 5.17, respectively), paired \(t = 2.81\) (df 3) \(p < 0.017\), one-tailed. This same finding replicated during problem-solving tasks (\(x\̄= 9.31\) v. 5.58, respectively), paired \(t = 6.89\) (df 3) \(p < 0.0015\), one-tailed. Overall, nonlinguistic fluency was elevated 56% during hypothesis generation when using the pen interfaces, compared with pencil and paper tools.

As shown in Figure 4B, analyses also confirmed the prediction that the keyboard-based interface increases students’ linguistic communication more than non-digital pencil and paper tools during hypothesis generation (\(x\̄= 32.17\) v. 46.42, respectively), paired \(t = 3.80\) (df 3) \(p < 0.008\), one-tailed. This finding replicated during problem-solving tasks (\(x\̄= 24.19\) v. 33.47, respectively), paired \(t = 5.80\) (df 3) \(p < 0.0025\), one-
tailed. Overall, linguistic fluency was elevated 41% during hypothesis generation when using a keyboard-and-mouse interface, compared with pencil and paper tools. In summary, these six analyses (see Table 2, A-H1-H6) all confirm that people communicate significantly more when using computer interfaces than a non-digital tool (PP). As detailed further in the following sections, the specific type of representations that are elevated during communication depend on the input capabilities of the digital tool (i.e., pen versus keyboard centric).

2.2.2. Pen Interfaces Prime More Nonlinguistic Fluency. To more fully confirm that pen interfaces increased students’ total nonlinguistic communication, the data were grouped and analyzed differently to reflect hypotheses summarized in Table 2. These further planned contrasts replicated that the pen interfaces (i.e., DP, PT) primed students to communicate significantly more nonlinguistic content when generating hypotheses than either of the other alternatives (PP, GT) ($\bar{x}$s = 7.0 DP and 8.0 PT, versus 5.2 PP and 5.3 GT), paired $t = 3.11$ (df 3) $p < 0.027$, one-tailed. This increased nonlinguistic fluency when using pen interfaces replicated again during convergent problem-solving tasks in session 3 ($\bar{x}$s = 8.9 DP and 9.7 PT, versus 5.7 PP and 7.9 GT), paired $t = 5.02$ (df 3) $p < 0.008$, one-tailed. It also replicated during convergent problem solving when nonlinguistic fluency was assessed as thinking marks, rather than as numbers, symbols and diagrams ($\bar{x}$s = 3.2 DP and 3.0 PT, versus 2.0 PP and 1.6 GT), paired $t = 3.18$ (df = 5), $p < .015$, one-tailed. In addition, nonlinguistic fluency was significantly more elevated when high-performing students used pen interfaces to generate hypotheses and solve problems ($\bar{x}$s = 7.3 DP and 7.2 PT, versus 6.2 PP and 4.8 GT), compared with the non-digital tool and keyboard interface alternatives, paired $t = 4.80$ (df, 3), $p < .009$, one-tailed. Total nonlinguistic fluency also was more elevated when low-performing students used the pen interfaces, compared with these other alternatives ($\bar{x}$s = 6.7 DP and 8.7 PT, versus 4.2 PP and 5.7 GT), paired $t = 4.64$ (df, 3), $p < .009$, one-tailed. In summary, these five analyses (see Table 2, B-H7-H11), combined with those reported in the last section (A-H3-H4), all replicate the finding that people communicate significantly more nonlinguistic content when using pen interfaces.

---

**Fig. 5.** Total hypothesis generation when using pen interfaces, compared with non-digital pencil and paper (A). Regression analysis showing positive relation between nonlinguistic communicative fluency and ideational fluency (B) [Note: Regression based on students for whom adequate nonlinguistic communication data was available, and removing far outliers]
2.2.3. Nonlinguistic Fluency Facilitates Scientific Ideation. In parallel, students produced significantly more appropriate biology hypotheses when using pen interfaces, compared with the non-digital pencil and paper tool, a significant difference by paired \( t = 3.44 \) (df 3) \( p < 0.011 \), one-tailed. Figure 5A illustrates that the total number of ideas generated was 38.5% higher for the same students working on the same tasks when they used the pen interfaces. This finding replicated when further analysis evaluated students’ hypothesis generation using the pen interfaces (DP, PT) compared with all other alternatives (PP, GT), \( \bar{x} = 6.9 \) DP and 7.5 PT, versus 5.2 PP and 5.7 GT, a significant difference by paired \( t = 4.47 \) (df 3) \( p < 0.01 \), one-tailed. No significant difference was observed between high- and low-performing students in this basic pattern, independent \( t = 1.88 \) (df, 14), \( p < .09 \). Regarding the number of hypotheses produced, high performers averaged 7.5 DP and 7.8 PT, versus 6.1 PP and 5.9 GT. The low performers averaged 6.3 DP and 7.2 PT, versus 4.3 PP and 5.5 GT.

A least squares regression analysis shown in Figure 5B reveals a significant positive predictive relation between low- and high-performing students’ nonlinguistic communicative fluency and their ideational fluency, \( F = 15.50 \) (df 1, 6), \( p < .008 \), \( R^2 \) of .72. In fact, knowledge of students’ nonlinguistic fluency while using different interfaces accounted for a substantial 72% of the variance observed in their idea production. The hypotheses for these planned analyses are listed in Table 2 (D-H17-H21).

2.2.3. Active Pen Marking Associated with Improved Problem Solving. All 16 students produced both diagrams and informal thinking marks while working on biology problems. As shown in Figure 6A, an analysis of matched pairs of problems revealed that when students constructed diagrams their scores averaged significantly higher than when they did not (\( x = 82\% \) v. 46%, respectively), paired \( t = 4.27 \) (df 19), \( p < .001 \), one-tailed. This difference represents a substantial 36% absolute improvement. As likewise shown in Figure 6B, when students made thinking marks on visuals displayed in problems before they started working on them, their scores also averaged significantly higher than when they did not (\( x = 89.5\% \) v. 65%, respectively), paired \( t = 2.37 \) (df 18), \( p < .015 \), one-tailed. This represents a substantial 24.5% absolute improvement. In these analyses, all comparisons were matched on the same problem, interface used, and student. The hypotheses for these planned contrasts are listed in Table 2 (D-H20-H21).

![Fig. 6. Scores on matched problems when students did or did not make a diagram (A), and when they did or did not make thinking marks on problem visuals (B).](image-url)
2.2.4 Keyboard Interfaces Prime More Linguistic Fluency. To more fully confirm that keyboard interfaces increased students’ total linguistic communication, the data were further analyzed according to the hypotheses listed in Table 2. These additional analyses replicated that linguistic communication was elevated significantly when students used the keyboard interface to generate hypotheses, compared with the pen interfaces or non-digital tools (\( \bar{x} = 43.8 \) GT, versus 32.0 PP, 33.4 DP, 31.6 PT), paired \( t = 2.78 \) (df 3), \( p < .007 \), one-tailed. This significant increase in linguistic fluency replicated when students used the keyboard interface during convergent problem-solving tasks in session 3, (\( \bar{x} = 33.5 \) GT, versus 24.2 PP, 25.7 DP, 25.0 PT), paired \( t = 3.96 \) (df = 3), \( p < .015 \), one-tailed. It also replicated in high-performing students during idea generation and problem-solving tasks, \( paired \ t = 6.04 \) (df 3), \( p < .005 \), one-tailed, and again in low-performing students during idea generation and problem-solving tasks, \( paired \ t = 5.59 \) (df 3), \( p < .006 \), one-tailed. In summary, these four analyses (see Table 2, C-H12-H15), combined with those demonstrated previously (A-H5-H6) all replicate the finding that people communicate more linguistic content when using keyboard interfaces.

Figure 7 illustrates that students also produced a significantly higher ratio of full sentences to clauses when using the keyboard-based interface, compared with other interfaces, Wilcoxon Signed Ranks test, \( T^+ = 59 \) (N 11), \( p < .01 \), one-tailed. Overall, the keyboard interface had a 56% higher ratio of sentential to clausal constructions than non-keyboard ones. A break-down of this ratio confirmed that students’ communication of full sentences increased significantly between the non-keyboard and keyboard interfaces (\( \bar{x} = 2.13 \) v. 3.38, respectively), \( T^+ = 116.6 \) (N = 15), \( p < .007 \), one-tailed. In addition, their communication of brief clauses decreased significantly between the non-keyboard and keyboard interfaces (\( \bar{x} = 2.75 \) v. 1.38, respectively), \( T^+ = 51.5 \) (N = 10), \( p < .007 \), one-tailed. See Table 2 (H-16) for the related hypothesis.

2.2.5 Keyboard-induced Linguistic Fluency Associated with Suppressed Scientific Ideation. A least squares regression analysis shown in Figure 8 confirmed that greater linguistic fluency was associated with significantly suppressed ideational fluency. In this case, a negative predictive relation occurred between students’ linguistic communicative fluency and their ideational fluency measured during hypothesis generation, \( F = 8.15 \) (df 1, 5), \( p < .036 \), \( R^2 \) of .62. The more linguistic content a student communicated, the fewer appropriate biology hypotheses they provided. Furthermore, knowledge of students’ linguistic fluency accounted for 62% of the variation in their suppressed
hypothesis generation. See Table 2 (E-H22) for the hypothesis related to this planned analysis.

2.3 Study One: Discussion

2.3.1. Can interfaces be designed that stimulate both communicative and ideational fluency? This research reveals that computer interfaces have affordances that can substantially facilitate or impede ideation and problem solving. In the case of education, interfaces can be designed that facilitate students’ own thinking and problem-solving activities—including increasing their communicative activity in a way that directly stimulates thought [Luria 1961; Vygotsky 1962]. This occurs because interfaces that facilitate communication involving representations central to a domain simultaneously stimulate and guide students’ mental effort associated with solving problems and constructing new schemas [Sweller, Van Merrienboer et al. 1998].

In accord with Affordance Theory, the results confirm that computer interfaces elicit more communicative fluency than non-digital pencil and paper tools. More specifically, interfaces characterized by different input capabilities, such as pen versus keyboard, have affordances that prime qualitatively different types of heightened communicative activity. Students expressed 56% more nonlinguistic representational content when using pen interfaces, compared with pencil and paper. More strikingly, they simultaneously generated 38.5% more appropriate biology hypotheses, and knowledge of students’ nonlinguistic communication level predicted 72% of the variance in their ability to generate appropriate ideas. In addition, expression of nonlinguistic diagrams and thinking marks was associated with 25-36% higher problem-solution scores, compared with matched problems in which students did no marking. These findings underscore the importance of designing future science interfaces that support active expression of nonlinguistic representations, including diagrams, symbols, numbers, and informal marks.

Although the GT interface provided free choice that included pen input, nonetheless its keyboard had dominant affordances that encouraged 41% more linguistic input than pencil and paper. Students also produced more full sentences when listing their hypotheses, which is consistent with a focus on presenting information to other people, rather than the task of generating new ideas. In sharp contrast with expression of nonlinguistic content, regression analyses revealed that
higher linguistic communication had a negative predictive relation with students’ idea production, accounting for –62% of the variance in their production of appropriate biology ideas. These findings are compatible with past literature showing increased linguistic input but degraded writing composition quality when using keyboard-based interfaces, compared with nondigital tools [Bangert-Drowns 1993; Haas 1989].

From a pragmatic viewpoint, these large magnitude differences in communication patterns indicate that digital pen interfaces have the potential to motivate and guide students’ problem-solving activities, especially in domains like STEM, more effectively than either keyboard-based graphical interfaces or non-digital tools. Clearly, future interfaces need to be designed to support students’ active spatial marking, no matter how informal it may appear. This includes diagramming, and also the meaningful placement of thinking marks within problem visuals as an aid in clarifying and self-organizing one’s work [Oviatt and Cohen 2010b]. In contrast, keyboard-based graphical interfaces are better suited for supporting interactive communication exchanges between students, or between a student and information resources.

2.3.2. What theoretical account explains interface facilitation of ideational fluency? From a theoretical viewpoint, human communication serves two functions: 1) a self-organizational aid to thought and performance, and (2) a means of conveying information to others and interacting socially with them. In this study, all students had extensive experience with keyboard-based graphical interfaces, which they mainly perceived as a tool for interacting with others and conveying information to them legibly. This view of keyboard-and-mouse graphical interfaces led to a higher rate of linguistic communication and full sentential constructions.

However, these same students had no previous experience or specific learned patterns associated with using pen interfaces. Affordance theory predicts that the physical similarity of pen interfaces with a pencil or pen would elicit perceptually-based expectations that their communication functionality is similar. This includes the broad coverage of pens for expressing different representations, and also marking and drawing as a self-organizational aid during thinking. Unlike non-digital tools, students’ understood that pen interfaces were computers. This may have motivated greater communicative activity, because the primary functionality of computers is serving as communication devices. These combined affordances could have increased students’ nonlinguistic communication when using the pen interfaces. Importantly, it was human perception of interface affordances that generated these differences in observed communication patterns, not simply the physical appearance or properties of the pen implements.

Consistent with Activity Theory, the present results also confirm that an interface that increases communication involving representations well matched with a domain can facilitate a parallel increase in ideation and problem solving. Activity theory emphasizes the important role of tools, including linguistic ones, in stimulating activity and guiding related thoughts [Luria 1961; Vygotsky 1962]. This study highlights that nonlinguistic representations, including spatial marking and diagramming, can be particularly valuable in mediating and refining mental activities, including producing appropriate ideas and solving problems correctly.

In contrast with the above account, a simple novelty effect motivating students to engage in generally higher activity levels cannot explain the fact that pen interfaces selectively stimulated more nonlinguistic communicative fluency, but not more linguistic fluency. It likewise does not explain why the keyboard-and-mouse interface with which students were most experienced stimulated the highest rate of linguistic communicative fluency. Finally, it is completely inadequate for explaining the specific nature of the strong coupling observed between heightened communicative fluency and ideational fluency, with nonlinguistic fluency predicting higher rates of ideation whereas linguistic fluency suppressed it.
Past computational research on input devices has been heavily biased toward evaluating their merit based on entry speed, or efficiency. This focus completely neglects the cognitive impact and benefits of any input device, and the fact that human communication serves the dual purposes outlined above. Furthermore, if human preference to optimize entry speed accounted for the present data, then students would have used keyboard-based interfaces to express clauses rather than full sentences, and they would have typed numeric input rather than writing it. Although keyboard entry often is assumed to be the fastest input alternative, students in this research actually completed the same biology tasks fastest when using the digital pen and paper interface, rather than either the pen tablet or keyboard-based graphical tablet [Oviatt and Cohen 2010a].

Future research needs to establish guidelines for when it is most beneficial to use different interfaces in educational contexts, and how to combine them most effectively with curriculum strategies that further enhance performance. The second study examines the generality of the present findings with: (1) inferential reasoning as a different type of cognition, and (2) a lower-performing student group. It also probes the important question of how interface tools that facilitate nonlinguistic communication are able to stimulate cognition so substantially.

3. STUDY TWO: THE IMPACT OF INTERFACE AFFORDANCES ON INFERENTIAL REASONING

The second study pursued a more specific exploration of whether a digital pen and paper interface significantly improves human cognition beyond that of analogous non-digital pen and paper tools. The type of cognition assessed was inferential reasoning, which is fundamental for all productive thinking. One goal in studying inference was to establish whether an interface’s input capabilities exert a broad influence on basic types of human cognition, including divergent ideation, convergent problem solving, and also inferential reasoning.

In pursuing this topic, the study examined whether a pen interface stimulates more fluent diagramming, along with a corresponding increase in accurate inferences about related content. It probed the specific dynamic relation between diagramming as a communicative activity and facilitation of correct domain inferences. For example, it explored the relation between multiple diagramming and reduction of overgeneralizations, the most common type of inference error. One potential advantage of promoting visual fluency skills, including appropriate diagramming and multiple diagramming, is that they can assist people with “seeing” how information should be scoped in terms of the breadth and definiteness of any conclusions.

This study also focused on lower-performing students, who potentially could benefit most from improved digital tools for thinking. Compared with existing keyboard-and-mouse interfaces, digital pen and paper interfaces have a variety of documented performance advantages for all students. However, these advantages are more extensive and larger in magnitude for lower-performing students [Oviatt, Arthur et al. 2006; Oviatt and Cohen 2010a]. This second study provided a combined intervention for low-performing students, which included: (1) an optimal interface tool (i.e., pens), and (2) a curriculum strategy that taught them how to construct a Venn diagram as a means of reducing overgeneralization errors, thereby increasing overall inference accuracy.

From a methodological standpoint, the second study also aimed to collect a larger amount of data on nonlinguistic representations, in this case diagramming, so that more precise analyses could be conducted on the dynamic relation between expression of nonlinguistic representations (diagrams) and facilitation of related inferences. To support this goal, the student population was increased in this study, and the Venn diagram intervention and instructions also encouraged students to diagram frequently.
3.1 Inferential Reasoning and Interface Design

People generate inferences in real-world contexts when they think about information, form impressions of other people, and engage in everyday decision-making. Inferential reasoning is a pervasive and fundamental cognitive skill, which is essential for all thinking and learning. It also is well documented that people have systematic biases in inferential reasoning [Kahneman, Slovic and Tversky 1982]. In particular, people have chronic difficulty with overgeneralizing information [Kahneman, Slovic et al. 1982]. For this reason, they frequently fail to qualify their conclusions appropriately in terms of breadth and definiteness of scope.

When people encounter situations involving high cognitive load in different contexts, this has been documented to reduce their inferences and increase inference errors [Kahneman, Slovic et al. 1982]. For example, recent research has documented that when students used a keyboard-and-mouse interface associated with higher cognitive load, their percentage of correct inferences decreased by 10%, compared with a lower-load pen interface [Oviatt, in press]. One important objective for future educational interfaces is to support accurate reasoning, including reducing systematic errors that undermine cognition.

3.2 Diagram-Facilitated Reasoning

Diagrams can facilitate inferences in a variety of ways. An apt diagram makes information more visually available to think about, especially if it is well matched with a domain and makes alternative possibilities visually explicit [Bauer and Johnson-Laird 1993]. In Bauer & Johnson-Laird’s (1993) research on deductive reasoning, people who used diagrams that made alternatives explicit responded significantly faster and made 28% more correct conclusions than those using verbal information. Diagrams facilitate visual comparison, and assist people with focusing on relevant evidence [Larkin and Simon 1987; Oviatt; in press; Schwartz and Heiser 2006; Suthers and Hundhausen 2003; Tversky and Suwa 2009; Zhang and Linn 2008]. They can suggest certain solutions while ruling out others, reduce working memory load, and by-pass the need to manipulate symbolic meanings [Bauer and Johnson-Laird 1993; Johnson-Laird 1999; Oviatt in press; Stieff and Raje 2010]. These advantages all are compatible with improving cognition and facilitating learning.

This study investigates the impact of interface design on students’ construction of spatial representations (i.e., Venn diagrams), and also the use of these diagrams to make accurate inferences about the domain information that is represented. Venn diagrams are well suited for inference tasks, because they permit translating verbal information into a visual form that encourages exploration, comparison, and minimization of cognitive load. Conclusions often can be drawn directly from the information provided in a Venn diagram. Venn diagram representations also have a canonical correct form, so they are conducive to developing clear metrics.

3.3 Theoretical Background, Goals and Predictions

The first study documented that computer interfaces elicit higher rates of communicative activity than non-digital tools, with pen interfaces selectively stimulating more nonlinguistic fluency and keyboard interfaces more linguistic fluency. These findings are compatible with Affordance theory, which maintains that different tools have affordances that encourage different levels and patterns of activity, but not others. The second study aims to replicate these findings, while focusing on diagrams as a form of spatial nonlinguistic representation. This study looks in more detail at the qualitative nature of diagramming, and how diagramming behavior relates to students’ ability to make correct inferences. It also specifically
evaluates differences between a digital pen and paper interface, compared with directly analogous non-digital pen and paper tools.

Consistent with Activity theory, study one also documented that pen interfaces that increase nonlinguistic communicative activity stimulate a parallel 38.5% increase in appropriate idea production (i.e., biology hypotheses). A linear regression confirmed that an individual’s level of nonlinguistic fluency predicted 72% of the total variance in their ideational fluency, which demonstrates the direct manner in which communicative activity mediates thought [Vygotsky 1962]. The second study focuses on how constructing spatial representations (i.e., Venn diagrams) stimulates, guides, and refines thought about related domain content. It examines whether a digital pen interface has affordances that will encourage people to diagram not only more actively, but also more correctly than analogous non-digital tools. It also assesses whether the digital pen interface will stimulate and guide a correspondingly higher percentage of correct domain inferences. Activity theory predicts that an interface tool should stimulate an improvement in thinking if it generates an increase in related activity as a mediating agent, but not if it does not. As a result, the digital pen interface would be expected to increase students’ correct inferences when their work is diagram-mediated, but not when it is verbally mediated alone.

The second study aims to determine the impact of using a digital pen and paper interface, compared with analogous non-digital pen and paper materials, on the accuracy of students’ inferential reasoning about science and everyday content. It also examines the dynamic relation between constructing one or more diagrams and facilitation of inference accuracy. The specific hypotheses examined include whether: (1) a digital pen interface will elicit more diagrams and more correct Venn diagrams than a non-digital pen, (2) a digital pen interface will stimulate a higher percentage of correct domain inferences than a non-digital pen when students diagram, but not when they complete tasks using verbal information alone, (3) construction of multiple versus single diagrams will suppress overgeneralization errors, improving the accuracy of domain inferences, (4) a non-digital pen will stimulate a higher

<table>
<thead>
<tr>
<th>Table 4. Definition of Terms &amp; Dependent Measures</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Inferential reasoning</strong> is the process of deriving conclusions that follow from information. In this research, information was expressed in verbal problem statements that people were given as “true.” They either made correct or incorrect inferential conclusions, which provided dependent measures for comparing different interfaces and instructional conditions in the present research.</td>
</tr>
<tr>
<td><strong>Conceptual Inference</strong> refers to an inference made about information in a content domain.</td>
</tr>
<tr>
<td><strong>Verbatim Inference</strong> refers to an inference extracted directly from a specific Venn diagram representation that a person made, independent of whether the diagram was correct or incorrect. This metric examined students' basic ability to understand and extract correct inferences, given a specific concrete visual, separate from their Venn diagramming skill.</td>
</tr>
<tr>
<td><strong>Venn diagram</strong> refers to a visual representation of information elements and their relation to one another as sets, which show different possible logical relations. In this research, information was expressed as verbal statements within a problem, which then were translated into a Venn diagram representation. Venn diagrams were scored as correct if they contained the correct number of elements, relation between elements, and labeling.</td>
</tr>
<tr>
<td><strong>Overgeneralization errors</strong> involve responses in which the scope or definiteness of a conclusion is too general to necessarily be true, given specific information in the problem statement (e.g., “All banana slugs are bright yellow” is too broad to conclude from problem 1 in Figure 9. The breadth of scope of the conclusion drawn is too inclusive to necessarily follow directly from information provided. In other cases, overly definite conclusions are made, rather than qualifying them as “could be” true.</td>
</tr>
<tr>
<td><strong>Undergeneralization errors</strong> involve responses in which the scope or definiteness of a conclusion is too narrow or restrictive, given information in the problem (e.g., “Some of the poisonous animals are not banana slugs” is too restrictive given the information provided in problem 1 of Figure 9). In this regard, undergeneralization errors are too uncertain, overly cautious, and restrictive.</td>
</tr>
<tr>
<td><strong>Scoping Error</strong> refers to an error in which the breadth or definiteness of the conclusion drawn about information is not accurately bounded. The inference made could either be too inclusive and general (i.e., overgeneralization), or too restrictive and limited (i.e., undergeneralization). In these cases, conclusions are improperly scoped or qualified.</td>
</tr>
</tbody>
</table>
percentage of verbatim correct inferences than a non-digital pen, or inferences that are correct given a specific diagram, independent of whether the diagram is correct or not (see Table 4 for terms), and (5) students will have limited awareness about which type of tool (digital, non-digital) best supports their ability to make correct inferences. These study predictions are associated with 8 specific hypotheses and planned contrasts, which are outlined in Table 5.

### 3.4 Study Two: Methods

Twenty-three undergraduate Communications students participated in the study as volunteers for course credit. Males and female participation was approximately equal. Low-performing students were recruited and encouraged to participate. All were fluent English speakers, expressed an interest in technology, and were expert users of graphical keyboard-and-mouse interfaces. None had used a digital pen interface previously.

![Problem 1](image)

**Figure 9. Sample inference task**

Students were asked to work on eight unique inference problems involving three statements apiece, as illustrated in Figure 9. The problems involved everyday reasoning and biology content. During each session, students viewed these problems for three minutes apiece on an electronic whiteboard at the front of the class. They completed four problems apiece using: (1) non-digital pen and paper materials (PP),
and (2) a digital pen and paper interface (DP) based on Anoto technology [Anoto 2011]. In this study, the digital pen was a Livescribe PulsePen [Livescribe 2011]. For both conditions, they did their work on the paper materials provided. These digital and non-digital paper materials were comparable in size and provided unlimited writing space.

Students worked independently during the session, which lasted about an hour. Before starting, they received instructions and practice with using the digital pen interface and its basic features, and any questions were answered. They were told that it could be used like a regular pen, except that it has a tiny camera at the tip that processes digital ink within the dot pattern on digital paper. They learned that it could interpret the content of their writing, or transmit it to another computer. When using the digital pen, students were told to cross out and rewrite content if they made an error, just like the non-digital pen. In both conditions, they were asked to focus on their work, and not worry if it appeared messy.

Figure 10 illustrates the Livescribe digital pen that was used in this study. It is similar to the Maxell pen used in study one in its ability to support writing with ink, just like a non-digital pen. Both of these digital pens also support vibro-tactile feedback, and are slightly chunkier than a non-digital pen. However, the Livescribe pen differs from the Maxell digital pen in several ways. For example, it is turned on with a button on the side, which is confirmed with illumination of the OLED visual display. In contrast, the Maxell pen turns on automatically when the cap is removed, which is confirmed with vibro-tactile feedback. In this study, students were asked to turn on the digital pen by pressing the side button, and to turn it off when done with the study. In addition, the Livescribe pen has multimodal input and output, and students were shown its vibratory, auditory, and visual feedback capabilities. However, students did not need to use the pen’s auditory capabilities, nor did they actually transmit or process their written input.

Figure 10 illustrates the Livescribe digital pen with its OLED visual display on the side.

The general instructions informed students that they would play a game in which they would be given true statements about everyday information or science content to read. They were instructed to think of as many possible valid conclusions as they could that follow directly from the statements, but that their conclusions should follow directly from and be logically compatible with the statements provided. They were told to avoid making irrelevant or incorrect conclusions, and were given examples of both correct and incorrect conclusions.

In addition, a Venn diagram tutorial gave students a brief presentation on how to translate verbal information in the inference tasks into a Venn diagram. This included making a diagram and labeling its parts correctly to reflect the number,
type, and relation among information elements in problem statements. Students were shown examples of how the inference problems could be diagrammed more than one correct way. They also were shown different correct conclusions could be made based on these alternative diagrams. They were encouraged to make multiple diagrams, and to use them to think of valid conclusions. On average, the inference tasks could be diagrammed twelve different ways.

In addition to using different tools, in one instructional condition students were asked not to make any diagrams while completing their inference tasks, and to summarize their answers as verbal statements (Verbal inference condition). In another condition, they were told to “make one or more Venn diagrams” of the problem statements first, and then to summarize their answers as verbal statements (Diagram inference condition). Before the diagram condition, students were given a Venn diagram tutorial. During the main study session, half of the problems were completed in the verbal condition, and half in the diagram condition.

Each session ended with a questionnaire. It was designed to confirm that the inference problems and Venn diagram tutorial were moderate in difficulty, and to assess prior experience using the digital pen interface. It also asked students to indicate whether they would prefer to use the digital or non-digital pen if they had to perform their absolute best on a high-stakes AP biology exam.

The research design aimed to accomplish a sensitive within-subject assessment of students’ ability to make correct inferences while using different tools (digital versus non-digital pen) during verbal- versus diagram-mediated reasoning. It specifically evaluated whether the digital pen interface provided a selective advantage during diagram-mediated reasoning, but not during verbally-mediated reasoning. The position of interfaces during testing, and the pairing of specific inference tasks with each interface, were counter-balanced across participants. The main analyses involved a priori planned comparisons of the impact of the two different tools on dependent measures involving diagram fluency and inference accuracy.

The study design was organized to investigate the dynamics of how and why the digital pen interface is able to stimulate students’ ability to make correct inferences. As a result, within-subject planned analyses probed whether the digital pen interface stimulated more frequent diagramming, and more correct construction of Venn diagrams during the diagram-mediated tasks. In addition, planned analyses evaluated whether there was a corresponding increase in students’ percentage of verbatim correct inferences, and in their percentage of conceptually correct inferences. A linear regression evaluated the relation between students’ frequency of diagramming and the suppression of overgeneralization inference errors. See Table 4 for definition of terms and dependent measures. Also see Table 5 for a summary of the eight hypotheses and planned contrasts that were tested and confirmed in the present study.

A trained researcher conducted independent second scoring on 13% of all inference responses. Average inter-rater reliability for the total number of inferences was 96%. In addition, inter-rater reliability for the total number of diagrams was 99%, and total correct Venn diagrams 96%. The reliability for correct verbatim inferences was 94%, total inference errors 87%, and total scoping errors (i.e., overgeneralizations, undergeneralizations) 81%.

Data were available for analysis on 176 inference problems, 88 for which diagrams were constructed and 88 based on verbal information alone. The data represented participation by moderately and very low-performing students, whose average percentage of correct inferences ranged between 21.1% and 78.6% (mean = 50%).
3.5 Study Two: Results

Figure 11. Total number of Venn diagrams, correct Venn diagrams, and ratio of correct Venn diagrams of the total when using the digital pen interface (DP), compared with the non-digital pen (PP).

3.5.1. Pen Interface Stimulates More Nonlinguistic Fluency than Analogous Non-digital Pen. Compared with using a non-digital pen, when students used the digital pen interface they drew significantly more Venn diagrams per problem, Wilcoxon signed ranks test, $x^*$ = 1.17 versus 1.25, $z = 1.76$, $N = 9$, $p < .04$, one-tailed. They also constructed significantly more correct Venn diagrams per problem, Wilcoxon signed ranks test, $x^*$ = 0.82 versus 1.03, $z = 2.16$, $N = 16$, $p < .016$, one-tailed. In addition, the percentage of correct Venn diagrams out of the total drawn was significantly higher when using the digital pen interface, Wilcoxon signed ranks test, 69% versus 83%, respectively, $z = 1.84$, $N = 15$, $p < .033$. Overall, students’ percentage of correct Venn diagrams was 14% higher when using the digital pen interface. Figure 11 summarizes these differences in diagramming due to the type of tool used. Table 5 lists the three hypotheses related to these planned contrasts (A/B H1-H3). These hypotheses correspond most closely with the A and B hypotheses listed in Table 2 for the first study, but they expand our understanding of these earlier results.

3.5.2. Nonlinguistic Fluency Facilitates More Accurate Inferences. Students’ percentage of verbatim correct inferences, or inferences that were correct based on the diagram representation they made whether the diagram was correct or not, averaged 89.2% when using the non-digital pen versus 95.2% with the digital pen interface, a significant difference by Wilcoxon signed ranks test, $z = 1.73$, $N = 12$, $p < .042$, one-tailed. Verbatim correct inferences distinguish students’ ability to make a correct inference from their ability to make a correct diagram.

As shown in Figure 12, students’ percentage of correct conceptual inferences during diagram-based inference tasks improved significantly from 38.3% when using the non-digital pen to 47.7% with the digital pen interface. According to predictions, when inference tasks involved construction of nonlinguistic communicative content (diagrams), the digital pen interface significantly stimulated students’ correct inferences, paired $t = 2.53$ (df =21), $p < .01$, one-tailed. This increase in correct conceptual inferences when using the digital pen was present in over 80% of the students sampled. This included the very low-performing group ($x^*$ = 25.2% v. 32.3%).
paired \( t = 2.34 \) (df = 11), \( p < .02 \), one-tailed. It also included the moderately low-performing group (\( x = 54\% \) v. 66.2%), paired \( t = 1.63 \) (df = 9), \( p < .069 \), one-tailed (marginal).

Most of the students in this study only made one diagram to represent information in each inference problem, although a subset of students did make multiple Venn diagrams for some or all of their tasks. For ten students there were adequate data to make a within-subject comparison of their percentage of correct inferences and inferential errors when they constructed one versus multiple diagrams. These students averaged 45.1\% correct inferences when they made multiple diagrams, compared with 35.8\% when they made a single diagram. This 9.3\% difference in correct inferences due to multiple diagramming was significant by Wilcoxon Signed ranks test, \( z = 1.89 \) (N = 10), \( p < .03 \), one-tailed.

Finally, since construction of multiple diagrams depicting different “possible worlds” can clarify the conclusions that potentially can and cannot be drawn about information, one specific hypothesis was that more diagramming would suppress scoping errors in inferential reasoning (i.e., overgeneralizations, undergeneralizations). To evaluate this prediction, a least squares regression was conducted. It confirmed that the number of Venn diagrams a student made predicted their percentage of scoping errors of total inference errors, \( R = .48, R^2 = .23, F = 6.39 \) (1, 21), \( p < .02 \). These results reveal that the number of diagrams predicted 23\% of the variance in students’ percentage of scoping errors, which were the most commonly occurring inference errors.

Table 5 lists the four hypotheses related to these planned contrasts (D H4-H7). These hypotheses correspond most closely with the D hypotheses listed in Table 2 for the first study. The present collection of results provides deeper insight into the interpretation of the previous findings.
3.5.3. Linguistic Fluency Does Not Facilitate More Accurate Inferences. As also predicted, the interface tools that students used did not have an impact on their performance during the verbally-mediated inference tasks. In this case, they averaged 57.5% and 55.5% correct inferences when using the non-digital pen compared with the digital pen interface, not a significant difference by paired *t* test, *t* < 1. Further examination of students’ total linguistic fluency during these verbally-mediated tasks also revealed no difference between the non-digital pen and digital pen interfaces (\(x^2 = 64\) v. 63, respectively), paired *t* < 1. Table 5 lists hypothesis E-H8, which motivated this planned contrast. It corresponds with the E hypothesis listed in Table 2 for the first study. In summary, it was confirmed that the digital pen interface selectively stimulated a 9.4% improvement in correct conceptual inferences in the diagram condition.

As shown in Figure 12, the low-performing students in this study performed better at producing correct inferences when using the non-digital tools if they did not have to do the additional work of first translating verbal information into a Venn diagram. When using a non-digital pen, they performed better when solely using verbal information as the basis for producing valid conclusions, \(x^2 = 57.2\%\) v. 38.3%, paired *t* = 2.78 (df = 21), *p* < .01, two-tailed. However, when using the digital pen interface, the correctness of their diagram-based inferences was no longer significantly lower than their verbally-mediated inferences, \(x^2 = 47.7\%\) v. 55.6%; paired *t* = 1.13 (df = 21), N.S. The use of the digital pen effectively reduced the gap between verbal-only and diagram-based inferences.

3.5.4. Students Believed Non-digital Pen Would Improve Performance. None of the students reported that they had previous experience using a digital pen and paper interface, although they all used it successfully during the session after a brief two or three minute orientation and practice. Paradoxically, when asked which type of pen they would prefer to use if they had to perform their absolute best on a high-stakes AP exam, 91% of students reported a preference to use the non-digital pen. Many said that it was more familiar and comfortable to use.

3.6 Study Two: Discussion

The present results are consistent with research described in the first study that computer interfaces support higher levels of communicative activity, with pen interfaces increasing nonlinguistic communicative activity in particular. Perhaps surprisingly, the digital pen interface facilitated substantially more accurate inferential reasoning, compared with an analogous non-digital pen. The results also elucidate the specific dynamics of how pen interfaces improved the accuracy of inferential reasoning. Basically, they stimulated a higher rate of diagramming activity, and more correct construction of diagrams representing domain information. Students’ construction of multiple diagrams also effectively suppressed the most common type of inference error, overgeneralizations. This study clearly underscores that diagramming is a critical type of nonlinguistic representation, and an important spatial foundation for human inference [Johnson-Laird 1999]. In addition, these results generalize the findings from study one by demonstrating pen interface enhancement of inferential reasoning, and facilitation in lower-performing students.

3.6.1 How do pen interfaces stimulate more accurate inferences? During inference tasks involving science and everyday reasoning, the digital pen interface stimulated significantly more total diagrams, more correct Venn diagrams, and a higher percentage of correct Venn diagrams of the total drawn. In fact, the percentage of correct Venn diagrams increased from 69% when using the non-digital pen to 83% with the digital pen interface. For the same students solving the same inference
problems, their percentage of correct Venn diagrams increased a substantial 14\% when using the digital pen interface.

Correspondingly, students’ percentage of correct inferences during diagram-based inference tasks improved from 38.3\% when using the non-digital pen to 47.7\% with the digital pen interface. When inference tasks involved diagram construction, the digital pen interface stimulated a higher rate of correct conceptual inferences in over 80\% of the students sampled, including both moderately- and very-low performing students. However, when inference tasks were completed using only verbal representations, the digital pen interface did not have an impact on students’ inference accuracy. In summary, the digital pen interface selectively stimulated a 9.4\% improvement in correct inferences, or the equivalent of one grade point.

To distinguish students’ basic ability to make a correct inference, given a correct diagram, from their competence at diagramming per se, additional analyses evaluated students’ percentage of verbatim correct inferences. Students’ ability to extract verbatim inferences increased significantly from 89.2\% correct when using the non-digital pen to 95.2\% with the digital pen interface. These results emphasize that students were very good at visually inspecting a diagram and extracting literal inferences. They also confirm the digital pen’s ability to facilitate this type of literal inference. Importantly, these data reveal that the digital pen interface significantly facilitated both: (1) correct diagramming of problems, and (2) correct reasoning about information represented in a diagram.

Table 6 summarizes the digital pen interface’s significant enhancement of correct Venn diagrams, correct verbatim diagram-based inferences, and correct domain inferences. The magnitude of enhancement was largest for construction of a correct diagrammatic representation (+14\%). When students had to both construct a diagram and use it to reason correctly about information within a domain, the magnitude of facilitation due to the digital pen interface still averaged a substantial 9.4\%.

<table>
<thead>
<tr>
<th>Type Representation or Reasoning</th>
<th>Pen</th>
<th>Digital Pen</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correct diagram</td>
<td>69%</td>
<td>83%</td>
<td>+14%</td>
</tr>
<tr>
<td>Correct verbatim diagram-based inference</td>
<td>89.2%</td>
<td>95.2%</td>
<td>+6%</td>
</tr>
<tr>
<td>Correct domain inference</td>
<td>38.3%</td>
<td>47.7%</td>
<td>+9.4%</td>
</tr>
</tbody>
</table>

Two related analyses further probed the dynamics of how pen interfaces improved the accuracy of inferential reasoning. They revealed that when students made multiple diagrams of a problem, they reduced their scoping errors. Scoping errors were by far the most common type, accounting for 78.9\% of all errors in this study. A linear regression confirmed that the number of diagrams constructed by a student predicted 23\% of all the variance in their percentage of scoping errors of the total. That is, more frequent diagramming was a significant predictor of suppression of scoping errors, such as overgeneralizations. In addition, a within-subject comparison revealed that students who made multiple diagrams averaged 45.1\% correct
inferences, compared with just 35.8% when they only made one diagram. The construction of multiple diagrams represents alternative “possible worlds,” which can clarify the conclusions that can and cannot be made about information. As a strategy, multiple diagramming supported students in thinking about different possibilities, which led to more appropriately qualified or scoped conclusions.

In spite of the relationship between diagramming and improved performance, most students only made one diagram per problem because diagramming is effortful. In addition, the three-minute time limit may have dissuaded some students from constructing extra diagrams. One problem with making a single diagram is that it encourages hyper-literalism in assuming that the diagram is the only correct answer, without qualification, and that no other “possible worlds” exist. This mental set represents a kind of rigidity, or functional fixedness, which is ripe for producing overgeneralization errors. In this study, low performers who only made one diagram actually performed better in the verbal-only condition, in part because their overgeneralization errors were extremely high when they made a single diagram. In contrast, using the digital pen interface stimulated more diagramming, even though it was effortful, which improved their performance to the level observed in the verbal-only condition.

3.6.2 What theory explains pen interface facilitation of inferential reasoning? From a theoretical viewpoint, these results can be viewed in terms of the synergistic interplay between Affordance and Activity theories. The digital pen and paper interface has affordances that elicited higher levels of nonlinguistic communicative activity, which in this study involved fluent diagramming. Students’ engagement in constructing diagrams created an opportunity to explore these visualizations, which invited actively thinking about the content represented. This stimulated and directly guided more accurate domain reasoning, including more refined scoping of conclusions [Vygotsky 1962]. The net impact was a chain of activity—ideation refinement, culminating in

Fig. 13. Chain of activity—ideation refinement, in which the digital pen interface encouraged more active participation

more correct reasoning about science and everyday information. Figure 13 illustrates this cascading sequence of events, in which the digital pen and paper interface encouraged more active participation by students. In this regard, the digital pen’s affordances invited certain activity patterns (i.e., constructing spatial representations), and the ensuing activities led to refinement of students’ ideation. Since the digital pen interface and its non-digital analogue were physically so similar, human perception of interface affordances played a major role in generating the activity patterns observed, not the physical attributes of the pen and paper alone.

From the viewpoint of Cognitive Load theory, the Venn diagrams that students constructed increased their germane load, or effort required to progress toward reasoning accurately about domain content. Diagramming was an effortful activity, which was clear from students’ questionnaire feedback. It also was clear from the fact that students typically only constructed one diagram per problem, even though they
were encouraged to make multiple diagrams. The digital pen interface effectively encouraged students to engage in a higher level of diagramming, in spite of the effort required. In this regard, it reduced the threshold for getting students to actively mark, draw, and begin the process of engaging in learning. This is a critical characteristic of an effective educational interface, so learners can become active participants in acquiring and automating new schemas [van Merrienboer and Sweller 2005].

A related issue is that a digital pen interface supports constructing spatial representations at different levels of sophistication. High-performing students may initiate the “chain of activity” by structuring a coherent diagram. However, other studies have shown that low-performing students often first become active making more elemental thinking marks (see Table 1 for terms; Oviatt and Cohen 2010a). Only afterwards do they progress to making more structured domain-specific diagrams. In fact, low performers produce thinking marks at double the rate of high performers, and for them the pen interface elicits a 101% higher rate of informal marking than other input alternatives (PP, GT) [Oviatt and Cohen 2010a]. In summary, these data emphasize that pen interface affordances can lower the threshold for encouraging low performers to become active participants in making spatial representations, which launches them along the “activity chain.”

The results outlined in this study are incompatible with the interpretation that students’ percentage of correct inferences improved when using the digital pen simply due to a placebo effect, or because they believed a computational tool would assist them more than a non-digital one. If this had been the case, students’ performance should have improved in both instructional conditions. Furthermore, 91% of students reported on the questionnaire that they preferred to use the non-digital pen, if they had to perform their best on a high-stakes AP test. This finding directly conflicts with attribution of a placebo effect.

For the same reason as in study one, a simple novelty effect also cannot explain the present pattern of results. If a novelty effect were present, it would have motivated students to engage in a generally higher activity level when using the more unfamiliar pen interface tool. However, the evidence indicates that students produced more nonlinguistic content (diagrams), but not more linguistic content, when using the digital pen interface compared with the non-digital pen. In addition, a novelty effect would have improved inferential reasoning when using the pen interface in all conditions. Instead, there was a selective facilitation of nonlinguistic communicative fluency, with more accurate reasoning only during the diagram-mediated tasks, but not verbally mediated ones. Finally, a novelty effect explanation is inadequate for explaining the specific qualitative nature of the direct coupling observed between increased diagramming and improved inference accuracy. For example, it does not explain how overgeneralization errors were reduced through the construction of multiple diagrams.

4. GENERAL DISCUSSION

4.1 What makes pen input such a good interface agent for conceptual change?

Pen interfaces provide a single focused input tool for fluently expressing both nonlinguistic and linguistic representations, and for shifting rapidly and flexibly among them without impeding thought. A pen implement that facilitates casting information in different representations supports perspective shifting in thinking about a problem, which provides traction for conceptual change. In the case of spatial representations (e.g., diagrams), pen interfaces also assist people with viewing information and retaining it in short-term memory while they think and derive solutions. Creating diagrams and thinking marks aids in grouping visually explicit
information that can facilitate efficient search, recognition, and reasoning about relational information as people extract inferences. In these important respects, pen interfaces that support spatial representations potentially provide a unique advantage over interfaces that focus on linguistic content [Oviatt, Arthur et al. 2007; Schwartz and Heiser, 2006].

The fertile interplay between pen input as a carrier of diverse mental representations and stimulation of cognitive flow during problem solving is fundamentally also what makes pen input valuable during design activities that require rapid generation and iteration of ideas. During both learning and design, pen interfaces support the exploration of possible meanings during communicative activity itself. Pen input invites exploring ideas during the physical process of marking or writing, which is characteristic of an effective tool for stimulating conceptual change. Educational psychologists have referred to communicative activity in the speech modality in an analogous vein, pointing out that talking can at times be an activity that explores possible meanings and precipitates insight, while at other times talking designates meaning:

“Talk is both an activity in itself, and an activity that makes other activities its topic.” —Roth [2005, p. 68]

Based on empirical findings, recent educational psychology and psycholinguistic literature also has argued that language activity and conceptual change co-emerge [Roth 2005], and that a given cultural group’s language directly influences its perception, cognition, and memory for information [Bloom, Peterson, Nadel and Garrett 1996; Levinson 2003]. Cross-cultural, developmental, and psycholinguistic research on these topics now reveals that the cognitive sciences previously underestimated the transformative power of language on human cognition [Bloom, Peterson et al. 1996; Levinson 2003]. The results from both studies described in this paper on communications interface design, including findings from a coherent set of 30 analyses, provide convergent evidence for this neo-Whorfian view of the impact of language on human cognition.

4.2 Research Strategy and Analyses

The present research investigated a new perspective on computer input capabilities as thinking tools, motivated by cognitive science theory and recent empirical results. It was unusual in the extent to which it aimed to comprehensively assess a series of specific targeted hypotheses related to the main research themes. The objective of this body of work was to assess the impact of computer input capabilities on human cognition, and to determine the generality of results across different types of basic cognition (idea production, problem solving, inference), domain content (biology, everyday common knowledge), student ages (high school, undergraduate), performance capabilities (low-performing, high–performing), and specific digital pens (Maxell, Livescribe). In order to systematically document the novel body of results presented here, the research strategy also focused on carefully assessing convergence in the pattern of results across different metrics to establish evidence for the research themes and their theoretical foundations.

Based on the themes summarized in Tables 2 and 5, thirty individual hypotheses were derived to conduct the present targeted examination described in studies one and two. Each hypothesis was a conceptual unit associated with an a priori planned contrast, which usually was directional and therefore assessed using a one-tailed probability value. The pattern of results that emerged was a substantial and coherent set of statistically significant findings, many large magnitude effects, which revealed a consistent pattern of support for the original research themes and their theoretical underpinnings.
Although many exploratory forays begin with a gunshot ANOVA and omnibus F test to investigate the possible presence of any research effect at all, such an approach would have been inappropriate for evaluating the specific directional hypotheses outlined in Tables 2 and 5. The selection of an appropriate statistical test should reflect the conceptual nature of the individual hypotheses formulated. An additional serious problem with the use of an ANOVA and follow-up tests in the present case would have been inadequate protection against type II inferential errors during statistical decision making, leading to overly conservative judgments and a failure to detect effects actually present in the data. When evaluating experimental data, type I and type II statistical inference errors are equally important to avoid, and protocols for accomplishing balance in achieving this goal have been debated among statisticians. Roger Kirk, a leading authority on statistics and the use of ANOVA in particular states:

“A significant F test indicates that something has happened in an experiment that has a small probability of happening by chance.” It allows you to reject the elementary but not very sophisticated “omnibus hypothesis of equality.” He further explains: “In planning an experiment, a researcher usually has in mind a specific set of hypotheses that the experiment is designed to test... called a priori or planned tests. This situation can be contrasted with... [one in which the researcher] believes that the treatment affects the dependent variable and the experiment is designed to accept or reject this notion.” If all “contrasts among means that are of interest to the researcher have been specified prior to collecting the data, multiple t statistics can be used to test the... hypotheses... it is not necessary to perform an omnibus test of the equality of the means before evaluating such contrasts.” —Kirk [1995 p. 113, 118, 172]

In summary, the research strategy adopted for documenting this substantial body of novel results included: (1) formulating targeted directional hypotheses based on theory and prior work, (2) confirming the generality of findings across different students, ability levels, domain content, types of cognition, and specific digital pens, and (3) evaluating the main research themes using convergent evidence based on many different metrics. The substantial collection of thirty findings reported here, including many large magnitude effects, provides coherent evidence beyond reasonable scientific doubt confirming the main research themes.

5. CONCLUSION AND FUTURE DIRECTIONS

The present results highlight an important direction for developing more expressively rich and flexible interface input capabilities, ones with affordances that can transparently but powerfully guide people’s ability to think, solve problems, and master science. They reveal that computer interfaces not only can motivate and engage people more than non-digital tools, but also increase their communicative fluency and facilitate related cognition. In other circumstances, basic computer input capabilities can substantially impede cognition, as demonstrated when students used keyboard input during science idea generation and problem solving tasks. The findings presented emphasize why pen interfaces are a promising direction for future educational interfaces. They elucidate precisely how pen interface affordances stimulate nonlinguistic communication patterns that mediate human cognition. They also demonstrate that the impact is a substantial and general one, which affects divergent idea generation, convergent problem solving, and inferential reasoning.

One important theme throughout this research is the role of interface affordances as facilitators of people’s own communicative activity. This design strategy has been demonstrated in past work with animated character interfaces, which successfully
elicited high-volume question asking (i.e., 150-300 questions in an hour) in 7-to-9-year-old students during marine science activities [Darves and Oviatt 2004]. It also has been demonstrated by interactive response systems in classrooms, which effectively prompted peer dialogue among undergraduates [Smith, Wood, Adams, Wieman, Knight, Guild and Su 2009]. The general design implications of this theme and related findings in this paper include that:

Interfaces play a major role in mediating and guiding basic human cognition; Their impact depends on affordances that stimulate increased communicative or physical activity, and on how well matched these activities are with a task domain.

Interfaces that support expressing different representations, especially nonlinguistic ones, provide greater expressive range, precision, and flexibility; They are capable of stimulating greater mental effort, and more refined and correct ideas.

The focus of these studies on science education provides a lens for discovering findings of more general import to computer interface design. STEM domains require frequent nonlinguistic information processing, which make them a good forcing function for examining the present topics. However, the expression of nonlinguistic representations is routinely required during real-world problem solving, especially whenever difficult phases arise. As a result, we might ask ourselves: How much of technology-mediated everyday cognition is impaired by keyboard-based input to computers? The present results challenge us to question the adequacy of existing keyboard-centric interfaces for future educational interfaces and also professional practice more broadly.

In the future, society could benefit from greater reliance on empirical evidence, which is capable of uncovering the real impact of technology design on human cognition. Perhaps surprisingly, people have very poor intuitions about the impact of computer interfaces on their own performance. This is demonstrated by the performance-preference paradox reported in the second study and other recent work [Oviatt, Arthur et al. 2006; Oviatt and Cohen 2010a]. Future research should explore the potential benefits of digital pen interfaces in situated learning contexts, over longer durations, and for lower performing and disabled students whose performance stands to benefit the most. In addition, it should examine the interface design principles addressed here in the context of fully functional systems, which may identify interactive design features that could further facilitate cognition. It also should investigate the most influential factors that determine people’s perception of interface affordances, which then motivate them to use an interface in specific ways. The development of a new generation of digital thinking tools is a long-term agenda, which will require intensive multidisciplinary partnerships between computational and learning scientists.

ACKNOWLEDGMENTS

Thanks to Inca Designs for financial support. Adrienne Cohen and Arianna Mann were supported by Inca Designs’ Science and Technology internships, and Andrea Miller and Kumi Hodge were supported by Inca Designs to attend an intensive one-week workshop on scientific/technical writing skills. We would like to thank to Russell Transue, Ali Maier and Kejun Xu for assistance with data collection and analysis, Alex Arthur for programming, and expert science teacher Ms. Korrie Beemer for valuable consultation regarding problem sets, scoring protocols, and related matters. Thanks also to Cliff Nass and Sean Westwood for generous assistance with arranging participants for study two. Research described in the second study originally was reported in briefer form at the International Conference on the Learning Sciences (ICLS) in Sydney Australia in 2012.

REFERENCES
