J Sci Educ Technol (2010) 19:515–531
DOI 10.1007/s10956-010-9218-7

Toward High-Performance Communications Interfaces for Science Problem Solving

Sharon L. Oviatt · Adrienne O. Cohen

Abstract From a theoretical viewpoint, educational interfaces that facilitate communicative actions involving representations central to a domain can maximize students' effort associated with constructing new schemas. In addition, interfaces that minimize working memory demands due to the interface per se, for example by mimicking existing non-digital work practice, can preserve students' attentional focus on their learning task. In this research, we asked the question: *What type of interface input capabilities provide best support for science problem solving in both low- and high-performing students?* High school students' ability to solve a diverse range of biology problems was compared over longitudinal sessions while they used: (1) hardcopy paper and pencil (2) a digital paper and pen interface (3) pen tablet interface, and (4) graphical tablet interface. Post-test evaluations revealed that time to solve problems, meta-cognitive control, solution correctness, and memory all were significantly enhanced when using the digital pen and paper interface, compared with tablet interfaces. The tangible pen and paper interface also was the only alternative that significantly facilitated skill acquisition in low-performing students. Paradoxically, all students nonetheless believed that the tablet interfaces provided best support for their performance, revealing a lack of self-awareness about how to use computational tools to best advantage. Implications are discussed for how pen interfaces can be optimized for future educational purposes, and for establishing technology fluency curricula to improve students' awareness of the impact of digital tools on their performance.

Keywords Educational interface tools and design · Biology problem solving · Pen interfaces · Digital paper and pen · Low-performing students · Meta-cognitive self-awareness

Introduction

New educational technologies have relatively sophisticated visualization and multimedia output capabilities for supporting STEM education (science, technology, engineering, mathematics) (Branford and Donovan 2005; Dede 2009; Linn et al. 2006; Wieman et al. 2008), although their user input capabilities remain far more limited. For educational interfaces to engage learners actively as they attempt to master content in simulations and other materials, this imbalance in interface design needs to be addressed. Existing keyboard-based graphical interfaces do not support students' active input well during the different phases of extended problem solving, which is a critical requirement for STEM education. Although graphical interfaces support linguistic and numeric input well, symbolic and diagrammatic representational content remain poorly supported.

Recent research has revealed that *basic computer input capabilities* can substantially facilitate or impede students' ability to produce ideas and solve problems in STEM domains. The presence or absence of a computer interface, its basic features, and the match of an interface to a task domain all play a demonstrable role in facilitating students' ideation and problem solving (Oviatt et al. 2010).

S. L. Oviatt (✉)
Incaa Designs, 11140 Wing Point Drive N.E.,
Bainbridge Island, WA 98110, USA
e-mail: oviatt@incaadesigns.org

A. O. Cohen
Duke University, Durham, NC, USA
e-mail: adrienne.cohen@duke.edu

Computer input features that facilitate students' communicative activity using representations central to an educational domain (e.g., diagramming for geometry) can maximize students' germane load, or effort associated with constructing new schemas (Oviatt et al. 2007, 2010; Oviatt and Cohen 2010; Sweller et al. 1998; van Merrienboer and Sweller 2005). As a separate but related theme, computer input features that minimize working memory demands due to the interface per se also play a role in facilitating ideation and problem solving. For example, computer input features that mimic non-digital work practice (e.g., digital pens) can leverage existing skills so students' attention remains focused on their learning task (Oviatt 2006; Oviatt et al. 2006). In summary, both maximization of germane cognitive load and minimization of extraneous load are two hallmarks of an effective educational interface.

Interfaces That Stimulate Both Communicative and Ideational Fluency in Science

Any high-quality educational interface must fundamentally be a rich communications interface, since communication is our most important tool for mediating learning. Educational interfaces ideally should be designed as tools that facilitate students' own engagement, communication, and problem-solving activities during learning—including increasing their communicative and physical activity in a way that directly stimulates learning. One reason that pen interfaces (e.g., digital paper and pen, pen tablet) are a promising candidate for facilitating students' active learning is that they encourage expressing a broad range of nonlinguistic representations, including informal marking and numeric, symbolic, and diagrammatic information (Oviatt et al. 2010). Pen interfaces also provide a single digital tool for easily shifting among different types of representation during the flow of working on solution steps. For example, when given a geometry word problem, a student may first diagram the relation between objects, then generate algebraic expressions using symbols and numbers, and finally summarize their calculations using linguistic content. Such a flexible flow of expressions can facilitate clarity of thought while working on extended STEM problem-solving tasks.

In recent research investigating science interfaces, the presence of a computer interface elicited more total communicative fluency than non-digital pencil and paper tools (Oviatt et al. 2010). When using computer interfaces to generate biology hypotheses, students' total communicative fluency was 15.5% higher than when they used hardcopy paper and pencil tools. This finding replicated when students worked on solving biology problems, with 23.5% higher total communicative fluency (Oviatt et al. 2010). In accord with Affordance Theory, this occurs because people primarily perceive computer interfaces as interactive communications tools, whereas pencil and paper are associated with noninteractive note-taking. *Affordance Theory* claims that people have perceptually-based expectations about interface tools, including constraints on successful performance that differentiate them. These interface affordances establish behavioral attunements that transparently but powerfully prime the likelihood that people will use them to communicate in certain ways (Gibson 1977). In summary, it is *human perception of interface affordances* that drives activity patterns when using a computer interface, not the presence of physical structures per se.

The same study revealed that interfaces characterized by different input capabilities, such as keyboard versus pen, have affordances that prime qualitatively different types of communicative action. Students expressed 44% more *nonlinguistic representational content* (numbers, symbols, diagrams) when using pen interfaces, compared with pencil and paper or a graphical interface that incorporated keyboard and pen input. This increase in nonlinguistic fluency when using pen interfaces replicated during hypothesis generation and problem-solving tasks, for low- and high-performing students, and when nonlinguistic fluency was assessed as "thinking marks" (i.e., informal pen marks produced on visuals displayed in problem statements) (Oviatt et al. 2010; Oviatt and Cohen 2010). More strikingly, students also generated 36% more appropriate biology hypotheses when using pen interfaces, and interface support for expressing nonlinguistic content predicted most of the variance in students' ability to generate appropriate scientific ideas (Oviatt et al. 2010). Figure 1 (left) illustrates that knowledge of students' nonlinguistic fluency while using different interfaces accounted for 72% of the variance in their appropriate biology hypothesis generation.

In contrast, a keyboard-based graphical interface stimulated 36% more *linguistic representational content* than either pen interfaces or pencil and paper. This finding was replicated during both scientific hypothesis generation and convergent problem-solving tasks, and for both low- and high-performing students (Oviatt et al. 2010). In this case, the regression analysis shown in Fig. 1 (right) revealed that increased linguistic communication was associated with undermining students' hypothesis generation, resulting in a *negative predictive relation*. Specifically, knowledge of students' linguistic fluency accounted for 62% of the variation in their hypothesis generation.

In addition to stimulating traditional forms of nonlinguistic communication, such as numbers, symbols and diagrams, pen interfaces facilitate more informal "thinking marks" than a graphical interface or pencil and paper (Oviatt and Cohen 2010). Thinking marks are informal pen marks that students make on existing problem visuals to help them

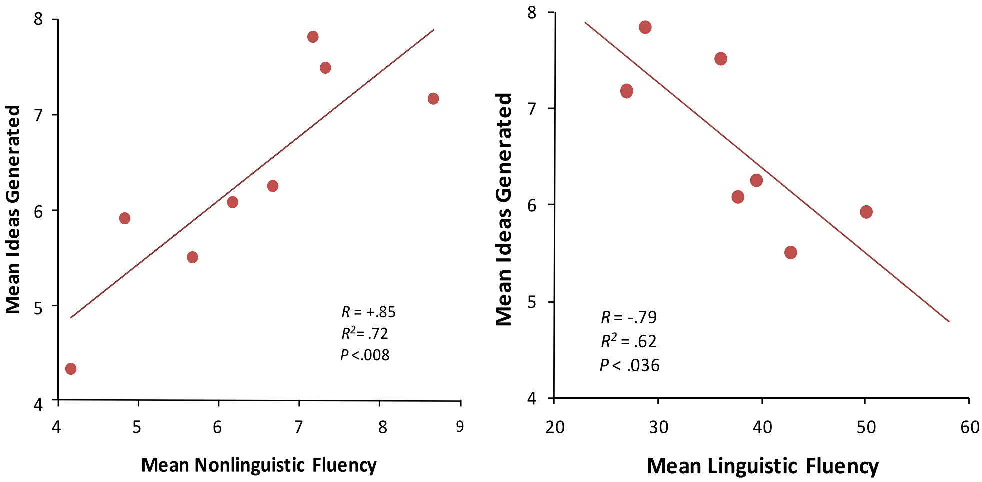


Fig. 1 Regression analysis confirming *positive predictive relation* between interface support for *nonlinguistic* communicative fluency and ideational fluency due to using a richly expressive pen interface (*left*); Regression confirming *negative predictive relation* between interface support for *linguistic* communicative fluency and ideational fluency when using a keyboard-based graphical interface (*right*)

count, select, order, group, label, and show relations between information as they work on solving a problem. This type of informal marking occurs at substantially higher rates in low-performing students, and has been associated with 24.5% higher science problem solution correctness (Oviatt et al. 2010; Oviatt and Cohen 2010). These data reveal that pen input that facilitates active marking may be valuable for scaffolding science learning, especially in lower-performing students (Oviatt and Cohen 2010).

Together, these findings underscore the importance of designing science interfaces that support active expression of nonlinguistic representations, including diagrams, symbols, numbers, and informal marks. They also confirm that an interface that increases communicative fluency involving representations well matched with a task domain can facilitate a parallel increase in ideational fluency and problem solving within that domain. These results are consistent with *Activity Theory*, which maintains that language plays a major role in mediating, guiding, and refining mental activities, such as problem solving (Luria 1961; Vygotsky 1962). From a functional perspective, for example, people increase self-talk spontaneously when they work on difficult tasks, and it is an effective strategy for guiding improved performance (Comblain 1994). The results outlined above on communications interface design are consistent with emerging neo-Whorfian views that a given cultural group's language directly influences its perception, cognition, and memory for information, which is based on a growing body of research in cross-cultural psycholinguistics, developmental psychology, educational psychology and related areas (Bloom et al. 1996; Levinson 2003; Roth 2005).

Interfaces that Minimize Load by Mimicking Existing Work Practice

In recent research comparing interfaces for mathematics education, a digital pen and paper interface was most effective at minimizing extraneous cognitive load (Oviatt 2006; Oviatt et al. 2006, 2007). A comparison of students' problem solution time, focus of attention, meta-cognitive control, correctness of problem solutions, and memory for problem content just solved revealed better performance when using a digital pen and paper interface than a pen tablet interface, which in turn supported better performance than a graphical tablet interface incorporating a keyboard, mouse and pen (Oviatt et al. 2006). Both high- and low-performing high school students took 16% longer to complete mathematics problem-solving tasks (i.e., introductory geometry and algebra) when using the tablet interfaces, compared with a digital pen and paper interface. In addition, all students' focus of attention deteriorated substantially when using the graphical tablet interface. Think-aloud protocols revealed that interface-related distractions when using the tablet interfaces, and especially the graphical tablet interface, undermined students' ability to maintain the high-level strategic thoughts needed to regulate their own performance. This was manifest as a significant reduction in the ratio of high-level think-aloud comments in which students specified what type of math problem they were working on, whether they were in an error state requiring repair, etc. from 17.3% when using the digital pen and paper interface, to 14.8% with the pen tablet, and 8.9% with the graphical tablet—or a 49% drop in high-level self-regulatory comments when using the graphical tablet compared with digital pen and paper interface (Oviatt et al. 2006).

In some cases, performance disadvantages of the tablet interfaces only were observed in lower-performing students. For example, lower-performing students' ability to remember the content they had just finished working on declined significantly from 69% when using a digital pen and paper interface to 61% when using the tablet interfaces (Oviatt et al. 2006). Their ability to solve problems correctly also was significantly more impaired when using the graphical tablet interface, compared with the pen interfaces. Specifically, correct math solutions by low-performing students dropped from 55% with the pen interfaces (i.e., digital pen and paper, pen tablet) to 39% with the graphical tablet interface. These latter results underscore that lower-performing students can experience new digital tools as a handicap when high-performing students do not, which is a circumstance that risks expanding the achievement gap between student groups during technology-supported learning activities.

Since this particular study focused on comparing the impact of *input capabilities* separately from system processing and output, one interpretation of this pattern of results is that the digital pen and paper interface minimized cognitive load most effectively because it most closely mimicked students' existing work practice during mathematics. In particular, it incorporated both a writing implement (i.e., pen input) and tangible paper medium. In comparison, the pen tablet interface included pen input but not the paper medium, and the graphical interface with a keyboard and screen least resembled students' existing mathematics work practice. From the standpoint of learning, a digital tool that mimics existing work practice minimizes students' extraneous cognitive load by leveraging already automated schemas for problem solving within a given domain. This, in turn, frees up mental resources to accelerate acquisition of new domain schemas.

These data highlight that designing interfaces to minimize cognitive load is especially consequential for low-performing students, who experience higher load than high performers when working on the same problems. It also is more consequential as problem difficulty increases overall. Compared with a digital pen and paper interface, tablet interfaces also include other sources of extraneous interface complexity associated with *system output*, including more frequent system interruptions and interface features (e.g., ink color and thickness formatting changes) that are unnecessary for learning tasks and therefore risk distracting students' attention (Oviatt 2006). These factors would be expected to further magnify any performance differences due to the interface beyond those reported.

Self-Awareness of Technology's Ability to Enhance Performance

If interfaces are developed that provide better support for problem solving and learning, the separate question that arises is whether students also will be aware of their beneficial impact and therefore prefer to use them. Although rarely investigated explicitly, a bias to believe in the performance-enhancing characteristics of technology has been documented in recent research. In the Texas laptop longitudinal study, both students and teachers believed that laptops in classrooms improved their learning and achievement, despite no significant improvement in students' self-directed learning or standardized test scores at any grade level in reading, writing, science, and social studies, and only small gains in mathematics at the 7th and 8th grade levels (Shapley et al. 2008; Shapley et al. 2009: Zucker and Light 2009). In research that evaluated deeper cognitive measures of performance, low-performing students believed that laptops would support them better than paper and pencil or a digital pen interface on an AP mathematics test. However, a controlled comparison revealed that students' mathematics scores actually dropped 11% when using the laptops, compared with either the digital pen and paper interface or hardcopy paper and pencil (Oviatt et al. 2006).

This latter finding is consistent with other research demonstrating that low-performing students with weaker meta-cognitive skills do not necessarily know how to organize and improve their own performance (Aleven and Koedinger 2000; Winne and Perry 2000), including judging what type of computing tools will best support them and when to use them (Aleven and Koedinger 2000). Research on self-directed help systems has indicated that students frequently do not have the skills needed to utilize such resources effectively (Aleven and Koedinger 2000). This is due primarily to more limited meta-cognitive skills in low-performing students (Aleven and Koedinger 2000; Oviatt et al. 2006; Winne and Perry 2000), which raises concerns that lower-performing students may be especially vulnerable to biases about the performance-enhancing impact of computer interfaces. In summary, even though students may be experienced users of computer interfaces, they do not necessarily have accurate self-awareness or judgment regarding how to use them to best advantage.

Related Literature on Pen Interfaces for Education

In recent years an extensive literature has emerged on pen interface design, some of which has focused on education and science content (Lajoie and Azevedo 2006; Pea and Maldonado 2006; Tabard et al. 2008; Yeh et al. 2006), nonlinguistic communication such as symbols, equations and drawings (LaViola and Zeleznik 2004; Oviatt et al. 2010; Tsandilas et al. 2009), and tangible paper-based interface design (Cohen and McGee 2004; Liao et al. 2005, 2008; Signer 2006; Yeh et al. 2006). Pen interfaces have many attractive features for education, including their expressive range (Oviatt et al. 2010), facilitation of informal work (e.g., sketching, brainstorming, design), and ability to bridge formal with informal learning environments (Cohen and McGee. 2004; Leapfrog 2009; Livescribe 2009; Oviatt et al. 2007, 2010; Pea and Maldonado 2006).

Tangible digital pen and paper interfaces, which span the physical and digital worlds, are beginning to be developed and commercialized (Adapx 2009; Anoto 2009; Leapfrog 2009; Livescribe 2009). They offer an ultra-lightweight and inexpensive interface that can be integrated with other devices such as interactive white-boards and laptops to create a flexible technology-rich learning environment. Digital pen interfaces are capable of cross-media integration of information sources, such as URLs and maps, which provides the opportunity to structure appropriately information-rich sources for supporting learning activities (Liao et al. 2005, 2008; Signer 2006; Yeh et al. 2006). In addition, permanent ink annotations and drawings on paper permit users to create spatial context, making information more mnemonic and easier to retrieve (e.g., concept maps), a significant advantage for study aids. Digital paper also has facile collaborative properties, including ease of sharing, copying, reorganizing, and reading or writing on it in different environments. Finally, recent research has contributed to the ethnography of digital paper and pen interface design (Bunt et al. 2009; Tsandilas et al. 2009), which in the future will require more research specifically within situated learning environments.

Study Goals and Hypotheses

The purpose of the present research was to investigate two questions:

- *What type of interface input capabilities provide best support for science problem solving in both low- and high-performing students?*
- *Are students aware of the impact of different computer interfaces on their performance, so they can benefit from them?*

To investigate these questions, the present study conducted a longitudinal comparison of high school students' problem-solving behavior while using alternative interfaces. Students completed biology problem-solving tasks during a series of three sessions in which they used: (1) pencil and paper materials, (2) an Anoto-based digital pen and paper interface (Anoto 2009), (3) a pen tablet interface, and (4) a graphical tablet interface that included keyboard, mouse, and pen. A comprehensive collection of performance measures was assessed, including problem solution time, self-regulatory behaviors, correctness of problem solutions, memory for problem content, and self-reported beliefs and preferences regarding different interfaces. These indices were used to compare how well different interfaces facilitated science problem solving, individual differences in interface facilitation of problem solving between low- and high-performing students, and any change over time in interface facilitation of problem solving between pre- and post-test sessions.

Since pen interfaces support expression of nonlinguistic representations that are central to science ideation and problem solving, it was predicted that the *same students solving the same problems* would experience greater performance facilitation when using the digital pen and pen tablet interfaces. Of these two, the digital pen and paper interface was predicted to support the best performance, since it most effectively minimizes cognitive load by mimicking existing work practice. The lower-performing students, who habitually experience higher cognitive load, were expected to benefit most from the digital pen and paper interface. All interface-related performance differences were predicted to persist over time, and to be evident in a convergent collection of cognitive indices. Finally, it was predicted that students would have limited self-awareness of the interface that actually best supported their performance, which has implications for both educational interface design and technology fluency curricula.

Methods

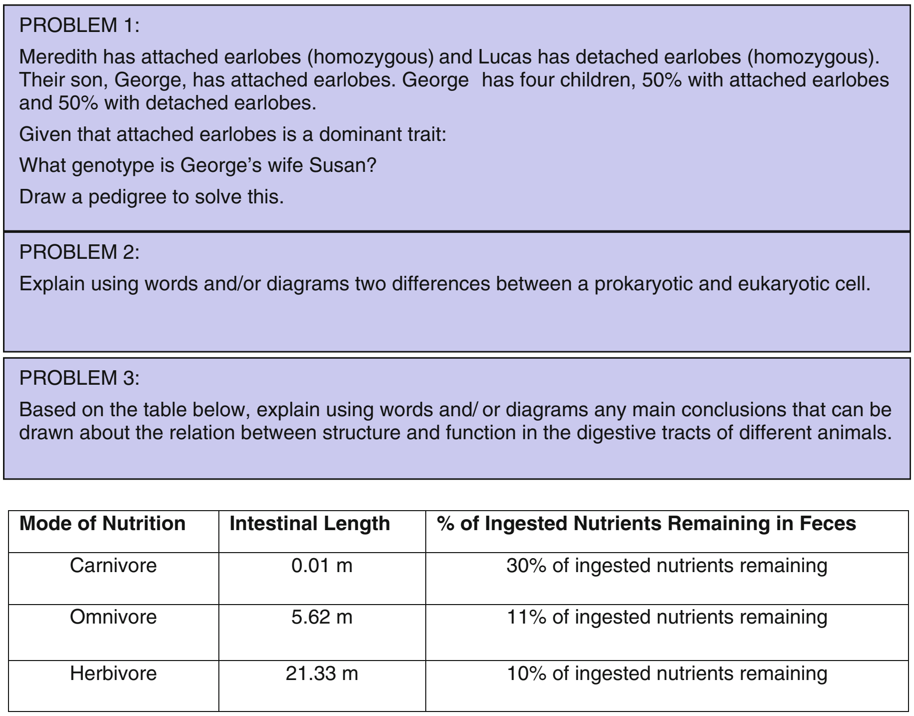
Participants

Sixteen public high school students who had recently completed an introductory biology class were included in the study as paid volunteers. 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 and also performance during the study. Within each of these subgroups, half of the students were female and half male. All students had been using non-digital tools in biology classes, but were experienced users of graphical interfaces with keyboard and mouse input for purposes such as text-editing, email, and information collection via the Web. All students also expressed an interest in technology, and were native English speakers.

Biology Problems

Based on consultation with biology teachers and advance pilot testing, eight parallel problem sets were composed, each containing four problems. These problem-solving tasks involved diverse content that students had just learned about in a year-long introductory biology class, such as genetic probability and inheritance, codon sequencing, cellular processing, physiology and ecological systems. Problems were selected that varied in both content and difficulty level to ensure a realistic range and generalizability of any results. Figure 2 illustrates three different

Fig. 2 Examples of biology problems



PROBLEM 1:

Meredith has attached earlobes (homozygous) and Lucas has detached earlobes (homozygous). Their son, George, has attached earlobes. George has four children, 50% with attached earlobes and 50% with detached earlobes.

Given that attached earlobes is a dominant trait:

What genotype is George's wife Susan?

Draw a pedigree to solve this.

PROBLEM 2:

Explain using words and/or diagrams two differences between a prokaryotic and eukaryotic cell.

PROBLEM 3:

Based on the table below, explain using words and/ or diagrams any main conclusions that can be drawn about the relation between structure and function in the digestive tracts of different animals.

Mode of Nutrition	Intestinal Length	% of Ingested Nutrients Remaining in Feces
Carnivore	0.01 m	30% of ingested nutrients remaining
Omnivore	5.62 m	11% of ingested nutrients remaining
Herbivore	21.33 m	10% of ingested nutrients remaining

types of sample problem. As illustrated, some of these introductory biology problems were more conceptual in nature, while others required mathematical computations or analysis of data summaries.

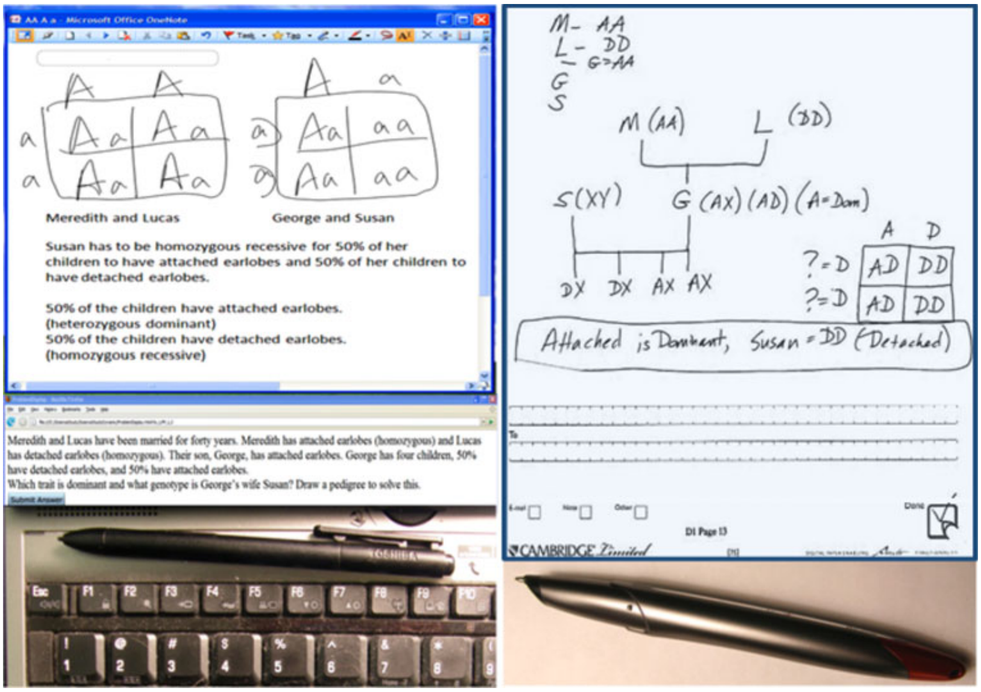
Interfaces and Equipment

During each session, students were asked to complete four problems apiece using: (1) non-digital paper and pencil materials (PP), (2) a digital paper and pen interface (DP) (Anoto 2009), (3) a pen tablet interface (PT), and (4) a graphical tablet interface (GT) that included a keyboard, mouse and pen. For all four conditions, each problem was presented on a Toshiba or Fujitsu laptop screen, as shown in Fig. 3.

When students completed a problem, they pressed a submit button on the tablet screen to view the next problem, which also logged auto-timing of completion time. In the two paper-based conditions students simply read the problem on the computer screen but did their work on paper. When using pencil and paper, they entered input with a pencil. With the digital paper interface, they wrote with a Maxell stylus on digital paper, as shown in Fig. 3. In the two tablet-based conditions, they entered their work directly on the computer using OneNote, a note-taking application. With the pen tablet interface, they used the tablet's stylus to enter all their input. When using the graphical tablet interface, they had free choice to use the keyboard, mouse, or stylus in any way they wished to enter input. Students were given comparable white space in all conditions to write or type their work. 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. In summary, students always had access to a writing implement in the four conditions. Table 1 summarizes the input and output features of all four conditions for comparison purposes.

Procedure

During the project's longitudinal sequence, students participated in three 1-to-1.5-h sessions that tracked their use of the different interfaces over time to ensure that any results would be durable. Students were given instructions and practice problems using each interface. For each of the three interfaces, students were shown the basic features required to complete 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 vibro-tactile feedback when: (1) removing the cap to start the pen computer, (2) making a checkmark in the "Done" box (Fig. 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. Students



Meredith and Lucas

George and Susan

Susan has to be homozygous recessive for 50% of her children to have attached earlobes and 50% of her children to have detached earlobes.

50% of the children have attached earlobes.
(heterozygous dominant)
50% of the children have detached earlobes.
(homozygous recessive)

Meredith and Lucas have been married for forty years. Meredith has attached earlobes (homozygous) and Lucas has detached earlobes (homozygous). Their son, George, has attached earlobes. George has four children, 50% have detached earlobes, and 50% have attached earlobes.
Which trait is dominant and what genotype is George's wife Susan? Draw a pedigree to solve this.

Fig. 3 Toshiba laptop screen with biology problem display on graphical interface, and keyboard, mouse and stylus for input (*left*, *bottom*); Student response using graphical interface with pen input and typing in OneNote (*left*, *top*); Student response using digital paper interface with Maxell pen (*right*)

Table 1 Summary of interfaces and their primary input, output and workspace features

Interface type	Problem statement[†]	Input tools	Work space
Pencil & paper (PP)	Tablet screen	Pencil	Hardcopy paper
Digital pen & paper (DP)	Tablet screen	Digital pen	Digital paper
Pen tablet (PT)	Tablet screen	Tablet stylus	Onenote on computer screen
Graphical tablet (GT)	Tablet screen	Keyboard, mouse & tablet stylus[‡]	Onenote on computer screen

[†] 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

were encouraged to work at their own pace, but to press the "submit" button as soon as they finished each problem. It was emphasized that they should just concentrate on solving each problem. However, if they couldn't complete a problem after trying, they should go on to the next. They practiced solving problems using each interface, after which they started the main test session once they had no more questions and were ready to begin.

During the pre- and post-test main sessions, each student worked individually while they completed 16 biology problems, four apiece using each interface. They were not given feedback on their problem solutions. Students also completed a questionnaire following their first and third test sessions in which they answered questions about their computer experience, perceived difficulty of the biology problems, what they liked and disliked about the different interfaces they had used, and how they would rank order the different interface materials on the following questions:

1. If you had to take an AP Biology exam and perform your best, how would you rank order your preference to use each interface or set of materials? (4 most preferred, 1 least preferred)
2. Which type of computer do you think is most likely to make errors?
3. Which type of computer is most likely to make your work easier?

After their third session, students were given a forced-choice memory test in which they were asked to recognize incidental information about the biology problems they had just finished solving while using the different interfaces.

Research Design

The primary goal of the research design was to accomplish sensitive within-subject testing of students' use of the four

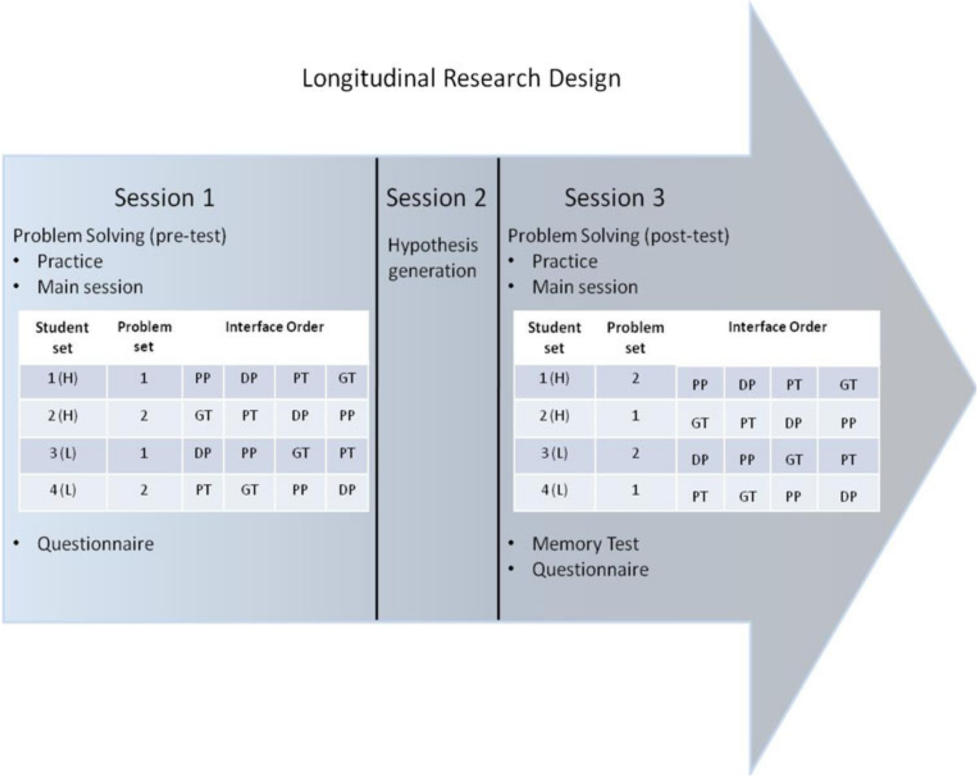


Fig. 4 Summary of procedural sequence and research design for pre-test and post-test sessions

Student set	Problem set	Interface Order			
1 (H)	1	PP	DP	PT	GT
2 (H)	2	GT	PT	DP	PP
3 (L)	1	DP	PP	GT	PT
4 (L)	2	PT	GT	PP	DP

Student set	Problem set	Interface Order			
1 (H)	2	PP	DP	PT	GT
2 (H)	1	GT	PT	DP	PP
3 (L)	2	DP	PP	GT	PT
4 (L)	1	PT	GT	PP	DP

types of interface during longitudinal testing. The order in which the four interfaces were presented was completely rotated within each group of four subjects. One of the two main problem sets was used for testing in session one and the other for session three, with counterbalancing between pre- and post-test sessions. Within the two main problem sets there were four problem subsets, each paired an equal number of times with the four interfaces during a test session. As such, each group of four students composed a completely-counterbalanced set for analysis purposes. Analyses of interface impact on the main dependent measures were conducted on the fully-counterbalanced subject sets, with interface order and specific problem sets considered nuisance variables. Analyses involving individual differences focused on low- versus high-performance status.

In summary, within-subject data were collected on the *same students solving the same biology problems*, so the study could provide a sensitive assessment of differences associated specifically with using alternative interfaces. The main analyses involved *a priori* planned comparisons of the impact of different interface types on student performance. Some analyses also compared pre- and post-testing of dependent measures, which reflected change or learning between the first and third sessions. The main features of the procedural flow and research design are summarized in Fig. 4.

Dependent Measures and Coding

Problem Completion Time

Timings were collected automatically from the beginning of each problem presentation until the student clicked "submit." Total time to solution then was computed for all problems that students solved correctly, excising cases where the solution was completely incorrect (i.e., score $= 0$) or not attempted at all. To assess replication of any effects, total time to solution also was rescored for all problem solutions receiving a score of 50% correct or higher. In both cases, total time was summarized as original raw data and logged data for each condition.

Problem Correctness

All biology problem-solving tasks from sessions 1 and 3 were scored for correct solutions out of 100%, and averages were summarized for each condition. Scoring templates, including protocols for giving partial credit, were developed based on consultation with an expert biology teacher.

Memory for Problem Content

The total percentage of forced-choice memory questions that students recognized correctly was summarized for each condition.

Self-Reported Questionnaires

Each student's rank-ordered preference for the four interface conditions was scored for select items on their written questionnaires (4 most preferred, 1 least preferred). Data also were summarized regarding student perception of the difficulty of biology problems, their experience using different types of computer interface, and what they liked and disliked about each interface. Average rank-ordered interface preferences were summarized, as was qualitative feedback about reasons for student preferences.

Metacognitive Self-Regulation

Two indices of meta-cognition were analyzed related to students' ability to make effective use of interface tools during science tasks. The first was self-awareness of the match between an interface's ability to support performance and students' stated preference to use the interface. For this purpose, data were compared on: (1) average percent correct solutions when using each interface, and (2) questionnaire rank-order preferences to use each interface for an AP biology exam.

A second meta-cognitive index was students' diagramming, a self-organizing strategy that has been associated with higher achievement on science tasks and improved self-regulation (Oviatt et al. 2010; Zhang and Linn 2008). The total number of diagrams (e.g., Punnett square) that each student produced while solving problems was averaged and summarized for each condition.

Reliability

Two expert scorers each independently scored all 512 biology tasks that students completed during sessions one and three for the correctness of problem solutions, with any scoring departures discussed and resolved so that 100% agreement was achieved. In addition, all diagrams were second-scored by an independent rater, with the average inter-rater reliability 100%.

Results

Data were available for analysis on 512 biology problems, 256 during session one and 256 in session three.

Problem Completion Time[1]

Individual Differences in Time to Complete

High- versus low-performing students did not differ significantly in average time required to complete problems during sessions one and three ($\bar{x}$ 213.0 vs. 197.0 s), independent $t < 1$, N.S.

Change in Completion Time

Figure 5 shows that the time it took for the *same students to solve the same biology problems* dropped substantially from the initial pre-test session to the final post-test. High-performing students sped up an average of 69.2 s, while low-performing students sped up 51.9 s, not a significant difference by paired t test, $t < 1$. The average time overall dropped from 242.1 to 181.2 s between the initial and final session, or 60.9 s total, a significant decrease by *paired t* test, $t = 4.75$ ($df = 3$), $p < .009$, one-tailed. This decrease represented a 25.2% drop by the final session.

Further examination of improvement in time required to complete problems between session one and three revealed that high-performing students averaged a larger drop when using the tablet interfaces than the paper-based ones ($\bar{x} = 85.5$ vs. 53.1 s, respectively), whereas the low-performing students improved more on the paper-based interfaces than the tablet ones ($\bar{x}$ drops 78.5 vs. 25.3 s, respectively). This contrast between high and low performers in interface facilitation of solving biology problems is illustrated in Fig. 6. Analysis of the ratio of improvement in time-to-complete scores when using the tablet versus paper interfaces confirmed that high-performers averaged significantly higher ratios than low-performers ($\bar{x} = 3.54$ vs. 0.62, respectively), independent $t = 1.98$ ($df = 13$), $p < .035$, one-tailed.

Table 2 summarizes the average decrease in time needed to solve problems while using the different interfaces for both high- and low-performing students. Individual analyses by paired t test further corroborate that time-to-complete scores were most facilitated in high performers when they used the tablet interfaces, although they were most facilitated in low performers when using paper interfaces. Table 1 also shows that the digital paper interface was the only one that significantly facilitated time improvement between sessions 1 and 3 for all students.

---

[1] Results reported in this section replicated for original raw data and logged data, and for analyses conducted on problems with scores above 0% or 50% correct.

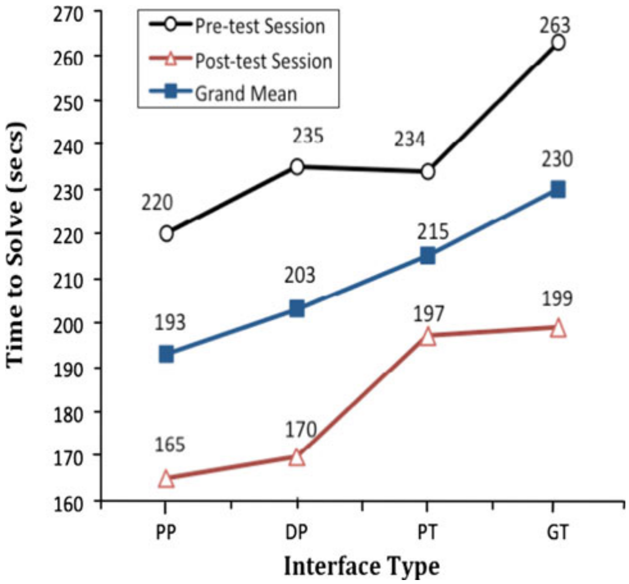


Fig. 5 Average time to complete problems using different interfaces during pre-test and post-test sessions

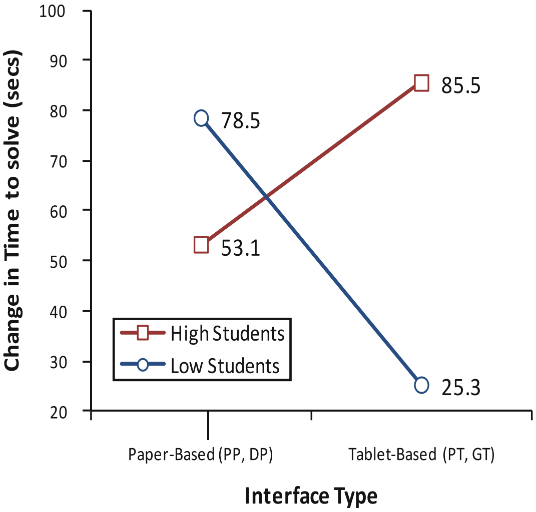


Fig. 6 Total improvement in time to complete problems between session 1 and 3 for high and low performers while using paper- versus tablet-based interfaces

Table 2 Improvement in time (seconds) to complete problems for high and low performers using different interfaces

Student	PP	DP	PT	GT
High performer	37.7	68.4**	81.9**	94.7**
Low performer	83.7**	73.3**	6.8	43.8

** Designates a significant decrease between session 1 and 3 at the $p < .05$ level

Interface Differences in Completion Time

Figure 5 shows the average time it took students to complete problems using different interfaces during the pre-test, post-test, and grand means. As predicted and illustrated, grand mean times increased as interfaces departed more from students' familiar work practice. The two paper-based interfaces did not differ significantly overall in time to complete, *paired* $t < 1$ ($df = 3$), N.S. Likewise, the two tablet interfaces did not differ overall in time to complete, *paired* $t < 1$ ($df = 3$), N.S. However, on the final post-test, paper-based interfaces averaged significantly faster than tablet-based ones, *paired* $t = 2.68$ ($df = 3$), $p < .04$, one-tailed. Further analysis confirmed that the digital paper interface was significantly faster than the pen tablet interface, paired $t = 3.62$ ($df = 3$), $p < .02$, one-tailed, and also faster than the graphical tablet interface, paired $t = 3.67$ ($df = 3$), $p < .02$, one-tailed. By the post-test, the pen and graphical tablet interfaces required 15.9% and 17.1% more time to solve problems, respectively, than using the digital paper interface.

Problem Correctness

Individual Differences in Problem Correctness

High-performing students averaged 78.2% correct on their problems in sessions one and three, compared with 57.4% for low-performing students, a significant difference by independent t test, $t = 4.64$ ($df = 14$), $p < .001$, one-tailed.

Change in Problem Correctness

The average score during session one pre-testing was 66.9%, compared with 67.1% during session three post-testing, not a significant difference by paired t test, $t < 1$, N.S. Neither the high-performing nor low-performing students' scores changed significantly between the first and third session, $t < 1$.

Interface Differences in Problem Correctness

Figure 7 shows average scores while using different interfaces during the pre-test session, post-test session, and grand means. As predicted and illustrated, the grand mean scores decreased as interfaces departed more from students' familiar work practice. Students' scores when using the paper-based interfaces were significantly higher overall than those when using the tablet-based interfaces ($\bar{x} = 72.0\%$ vs. 63.4%, respectively), paired $t = 5.93$ ($df = 3$), $p < .005$, one-tailed.

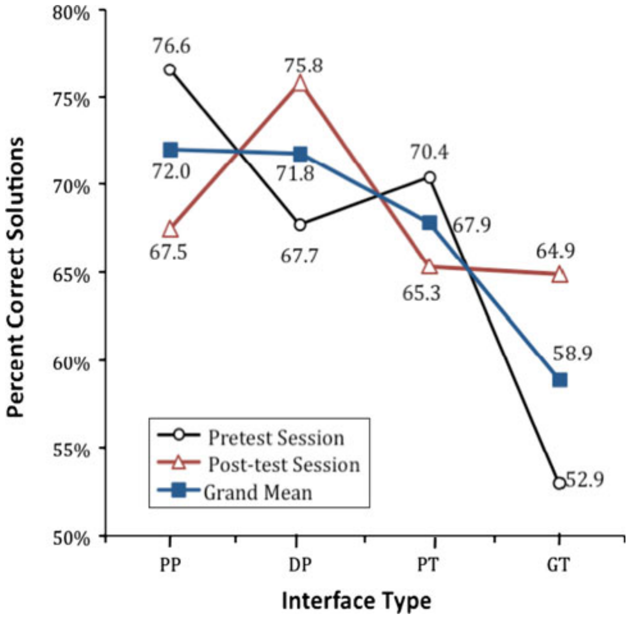


Fig. 7 Average percent correct scores using different interfaces on pre-test and post-test sessions

Students' performance when using hardcopy pencil and paper versus the digital pen interface did not differ significantly on the post-test, paired $t = 1.33$ ($df = 3$), N.S., and their performance when using the pen versus graphical tablet interface did not differ, paired $t < 1$, N.S. However, students performed better when using the digital pen interface ($\bar{x} = 75.8\%$) than the tablet interfaces ($\bar{x} = 65.1\%$), paired $t = 2.96$ ($df = 3$), $p < .03$, one-tailed. By the final post-test, scores when using the digital paper interface were 10.5% and 10.9% higher than when using the pen or graphical tablet interfaces, respectively.

Memory for Problem Content

Individual Differences in Memory

The percentage of content students recognized correctly for problems they had just completed during the final session averaged 61.5% for high-performing students and 63.3% for low-performing ones, not a significant difference by paired t test, $t < 1$, N.S.

Interface Differences in Memory

Students' memory for problem content did not differ significantly when using hardcopy paper and pencil materials versus the digital pen interface, paired $t < 1$, nor did it differ significantly when using the pen tablet versus graphical tablet interface, paired $t < 1$. However, as predicted, students recognized significantly more content after using the paper-based interfaces than the tablet-based ones ($\bar{x} = 65.9\%$ vs. 58.9%, respectively), paired $t = 9.01$ ($df = 3$), $p < .001$, one-tailed. As illustrated in Fig. 8, low-performing students' ability to recognize biology content dropped from 67.2 to 59.4% between the paper and tablet interfaces, a 7.8% absolute reduction. Likewise, high-performing students' recognition decreased from 64.6 to 58.3% between the paper and tablet interfaces, a 6.3% reduction. Overall, this represented a substantial 7.1% reduction in students' ability to remember biology content after using the tablet interfaces.

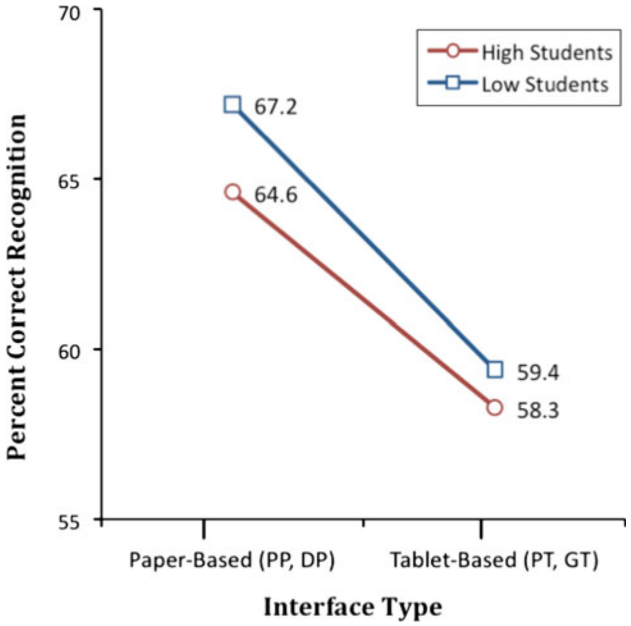


Fig. 8 High- and low-performing students' average percent correct recognition memory for content just solved after using different interfaces on post-test session

Self-Report Questionnaires

Problem Difficulty Ratings

Biology problem difficulty ratings averaged 3.25 for low-performers and 2.94 for high-performers on a 5-point scale, with no significant difference between groups in perceived difficulty. Both groups viewed the problems as intermediate in difficulty.

Experience Using Different Interfaces

In biology classes, students used non-digital pencil or pen and paper for most of their work. With respect to familiarity and expertise using different interfaces, all 16 students reported extensive multi-functional experience with graphical interfaces. Primary use of graphical interfaces involved text processing (e.g., including preparing biology lab reports), email/communications, and Web searching.

Only six students had ever used a pen tablet computer before. In these cases, they had either used it just once, or for a limited purpose (e.g., drawing). None of the students had ever used a digital paper and pen interface. High- and low-performing students did not differ in self-reported computer experience.

Interface Differences in Supporting Test Performance

When asked on session 3 which interface they would prefer to use if they had to perform their best on an AP Biology exam, Fig. 9 shows that high- and low-performing students rank ordered their preferences similarly—pen tablet (PT), graphical tablet (GT), paper and pencil (PP), followed by digital pen (DP). Analysis by Wilcoxon Signed Ranks test revealed that students preferred the tablet over paper interfaces, $T+ = 62$ ($N = 11$), $p < .007$, two-tailed. Further analysis of change between session 1 and 3 in students' preference to use different interfaces for an AP exam revealed accentuation over longitudinal sessions in their initial preferences. Figure 10 illustrates that there was a significant increase in preference to use tablet interfaces by session 3, and a decrease in preference to use the digital pen interface, Wilcoxon Signed Ranks, $T+ = 64$ ($N = 11$), $p < .003$, two-tailed.

Interface Differences in Making Work Easier

When asked on session 3 which type of computer was most likely to make their work easier, students' average rating for both the pen and graphical tablet interface was 2.25, compared with 1.5 for the digital pen interface. In fact, nine out of ten students who reported a clear preference rated the tablet interfaces higher than the digital pen interface, a significant difference by Wilcoxon Signed Ranks test, $T+ = 49.5$ ($N = 10$), $p < .025$, two-tailed.

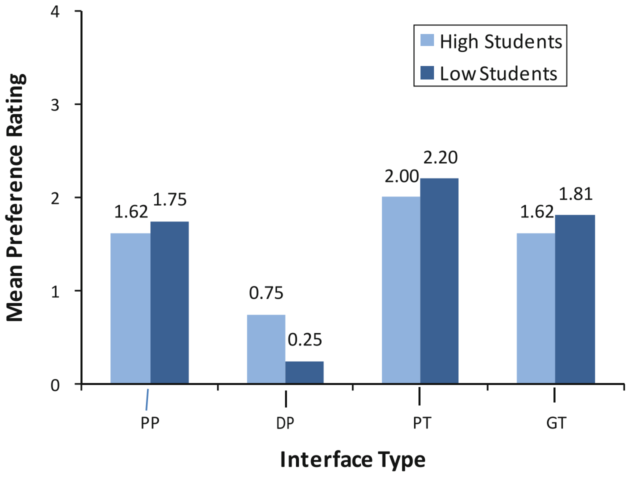


Fig. 9 Self-reported preference rating by high and low performers to use different interfaces for AP exam

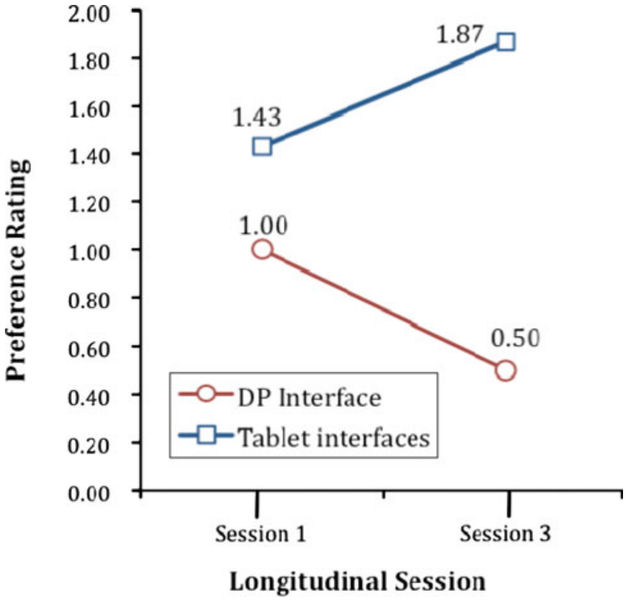


Fig. 10 Change between session 1 and 3 in preference to use digital paper versus tablet interfaces for AP biology exam

Interface Differences in Error-Proneness

When asked on session 3 which type of interface was most likely to generate errors, low-performing students did not rate the interfaces significantly differently ($\bar{x} = 1.93$ for digital pen vs. 2.04 for pen and graphical tablets). However, high-performing students rated the digital pen interface as significantly more error-prone than the tablet interfaces ($\bar{x} = 2.75$ for digital pen, 1.94 pen tablet, 1.31 graphical tablet), a significant difference by Wilcoxon Signed Ranks test, $T+ = 28$ ($N = 7$), $p < .02$, two-tailed.

Interface Features Most Liked and Disliked

The pen tablet was the preferred interface. It was considered fun, entertaining, and generally easy to use. Students loved being able to change colors, move ink objects around, and create extra space. However, their comments also revealed that writing on the tablet was viewed as awkward, slippery, uncomfortable, and not like writing on paper. Students thought the quality of their handwriting appeared less clear and messier. In addition, they disliked scrolling. In these respects, the pen tablet was fun but also distracting, which undermined students' focus of attention while they worked.

The graphical tablet interface was viewed as fast, familiar, and "organized" by many students, most of whom viewed themselves as good typists. They also liked the fact that they could draw with the pen. However, students commented that the graphical interface involved too many choices and too much coordination to use, which could be confusing. For example, they disliked the burden of switching between typing and pen input. They also disliked the awkwardness of their handwriting quality and writing on the tablet's surface.

The digital pen interface was considered simple, familiar, and easy to use while focusing on work. However, students disliked its lack of clear confirmations. They disliked not being able to erase. When crossing out ink, they were uncertain whether the computer recognized that their information had been deleted. They also would have preferred that their written input and computer feedback be in one integrated location.

Metacognitive Self-Regulation

Preference-Performance Paradox

Table 3 shows that for the 13 of 16 students who indicated a clear interface preference on session 3, 100% preferred to use the tablet interfaces for an AP exam, rather than the digital pen interface. This preference conflicted with an *11% average lower score when the same students used the tablet interfaces,* compared with their performance using the digital pen interface.

Advance Diagramming

All 16 students produced diagrams while solving some of their biology problems. Both low- and high-performing students produced the same number of total diagrams ($\bar{x} = .59$), with no significant difference between groups, $t < 1$, N.S. As shown in Fig. 11, the average number of diagrams students produced in sessions one and three decreased as interfaces departed from existing work practice. As predicted, students constructed diagrams significantly more frequently when using paper-based than tablet-based interfaces, paired $t = 15.69$ ($df = 3$), $p < .001$, one-tailed. In fact, the rate of diagramming when using the digital paper and pen interface averaged 31.4% higher than when students used the tablet interfaces.

Table 3 Self-reported preference to use digital paper versus tablet interfaces, compared with solution scores supported by these interfaces

	Digital pen interface	Tablet interfaces
% Problems Correct	76%	65%
Reported Preference	0%	100%

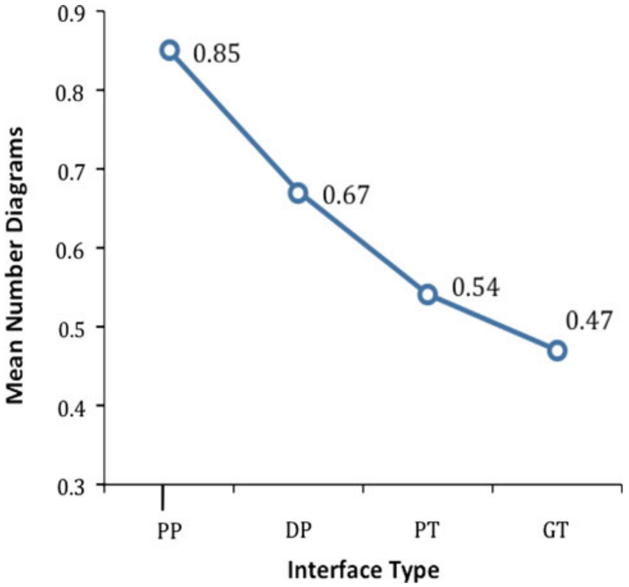


Fig. 11 Average number of diagrams produced when using different interfaces during sessions 1 and 3

Discussion

What Type of Interface Input Capabilities Provide Best Support for Science Problem Solving?

The tangible paper and pen interface supported significantly higher correctness of problem solutions than the tablet interfaces, and it did not degrade performance in comparison with existing pencil and paper work practice. On the final session, students' problem solutions averaged a substantial 10.5% higher when using the digital paper and pen interface than the pen tablet, and 10.9% higher compared with using the graphical tablet interface. Students also averaged significantly faster times for completing their problems when using the digital paper and pen interface, compared with the tablet interfaces. By the final session, students required 15.9% less time when using the digital paper interface than the pen tablet, and 17.1% less time than the graphical tablet interface. Similarly, by the final session students' memory for biology content they had just finished working on averaged 7.1% higher after using the digital paper interface, compared with the tablet interfaces. These converging results for science problem solving corroborate and generalize past findings obtained in the domain of mathematics, and confirm their durability after more extended longitudinal testing.

With respect to indices of self-regulation, active diagramming (e.g., Punnett squares for genetics problems) occurred at a significantly higher rate in the paper-based interfaces than tablet-based ones. In particular, the digital paper and pen interface facilitated a 31.4% higher rate of diagramming, compared with the tablet interfaces. This is especially noteworthy since related work has shown that diagramming is a self-organizing strategy associated with 36% higher solution correctness on biology problems (Oviatt et al. 2010), and better integrated comprehension of science content displayed in simulations (Zhang and Linn 2008). Pen interfaces that are perceived by students as facilitating diagramming and informal marking are especially important as problem difficulty increases and students' rate of active marking spontaneously rises (Oviatt et al. 2007), and also for lower-performing students who habitually experience greater cognitive load. In this study, all students produced diagrams as an aid to performance while solving problems, and both low- and high-performing students diagrammed a comparable amount.

Do High- and Low-Performing Students Benefit Equally from the Same Computer Interfaces?

Although the average correctness of problem solutions remained stable across sessions, nonetheless the time required for students to complete problems sped up 25.2% (i.e., from four to three minutes) by session three as their skills became automated during the process of schema acquisition. All students acquired faster problem-solving skills using all sets of materials, whether hardcopy or digital tools. However, lower-performing students' skills sped up significantly faster when using the paper-based tools (+210%), whereas the higher-performing students' skills consolidated significantly faster when using the tablet interfaces (+61%). Figure 6 illustrates this interaction in interface facilitation of solution times for different student groups. A comparison of the interfaces shown in Table 2 reveals that only the digital paper and pen interface facilitated significantly faster solution times for all students. In this regard, low-performing students benefited most from using the digital paper and pen interface, but both groups were well supported by this interface. In comparison, tablet-based interfaces impeded low performers' progress.

Previous research also has shown that low-performing students' solution correctness and memory during mathematics problem solving deteriorates relatively more than high performers when using tablet-based interfaces (Oviatt et al. 2006). These convergent findings across studies indicate that adoption of tablet-based interfaces for STEM problem-solving risks expanding rather than minimizing the achievement gap between student groups. Further research will be required to determine how interface facilitation of schema acquisition in low- versus high-performing students may continue to change over longer periods of exposure and in situated classroom contexts.

Are Students Aware of How Technology Influences Their Performance?

One index of meta-cognitive skill and ability to self-regulate one's learning is an accurate awareness of when and how to use digital tools to best support performance. Previous research involving mathematics revealed a *performance-preference paradox* for low-performing students. They reported a preference to use tablet interfaces during a high-stakes exam when they had to perform their best, even though their work was substantially worse when using tablet interfaces compared with a digital paper and pen interface (Oviatt et al. 2006). In the present research, this performance-preference paradox was replicated in both high- and low-performing students. Furthermore, evidence revealed that it became increasingly accentuated over sessions. When asked which interface they would use to perform their best on an AP biology exam, 100% of students expressed a preference for the tablet interfaces over the digital paper and pen, in spite of the fact that their solution correctness dropped 11% when using tablet interfaces. Although teenagers often are perceived as expert in operating computers, this disparity nonetheless reveals that their awareness of the actual impact of computers is limited. Surprisingly, this meta-cognitive limitation involved overconfidence in the graphical tablet interface, with which students had the most extensive experience. These illusions about the performance-enhancing capabilities of tablet interfaces appear to stem from students' self-reported beliefs that tablet interfaces make fewer errors and make their work easier.

As technology increasingly permeates schools and our lives, one major concern is the extent to which students and society in general are vulnerable to illusions regarding the perceived benefits of computing and its ability to compensate for their performance limitations. During critical periods of learning opportunity, if students adopt computing tools without using good judgment as reflected in realistic self-regulatory skills, then they risk chronically under-performing on many important tasks. Given the higher cognitive load experienced by low-performing students, poor meta-cognitive skills in guiding technology use can be expected to have the most damaging impact in the least capable students. To address these concerns, future technology fluency curricula will need to replace widespread illusions of technology as *knowledge dispenser* with a more reality-based model of computers as simply tools that:

- Do not lessen a learner's need to expend effort in order to acquire knowledge, or in any sense automate the learner's internal process of actively consolidating an understanding of the world

- Can facilitate one's own active engagement, communication, and problem solving during learning activities, but only if used under circumstances in which these tools do not overload the learner or distract from the main learning activity
- Can also undermine one's ability to perform well and learn, if interface affordances and features are poorly matched with a learning task

This instruction could include demonstrations in which students select a computer tool for a particular task, and then receive feedback on their performance outcome compared with using other alternatives. Follow-up discussion could focus on uncovering systematic biases in students' beliefs about what different computer tools actually can and cannot be used to accomplish. In addition to near-term curriculum issues related to teaching students more effective "technology fluency," education leaders will need to assume a far more proactive role in *defining specifications and driving the improvement of educational interface design*. It is simply not realistic to assume that computationalists and engineers are adequately knowledgeable to design optimal educational interfaces.

Why are Pen Interface Tools Well Suited for Stimulating Conceptual Change?

Pen interfaces, including digital paper and pen and the pen tablet, are effective at facilitating students' active problem solving and learning because they support expression of a broad range of linguistic and nonlinguistic representational content (Oviatt et al. 2010). They also provide a single low-load digital tool for easily shifting among different types of representation during the flow of working on solution steps, which facilitates clarity of thought while solving science problems. More specifically, pen interfaces support:

- Grouping of visually explicit information, which facilitates efficient search, recognition, and reasoning about relational information as people construct ideas (Larkin and Simon 1987)
- Viewing and retaining information in short-term memory while insights are derived
- More elaborated and precise learning (e.g., through diagramming objects), which deepens learning and improves transfer (Mestre 2005; Schwartz and Heiser 2006; Schwartz et al. 2008)
- Improved robustness of memory, due to dual coding in visual and verbal forms (Baddeley 1986; Paivio 1986)
- Figure-ground perspective shifts (e.g., by creating visuals, deriving calculations), which facilitate more frequent and novel insights (Oviatt et al. 2010)
- More fluent idea generation (Oviatt et al. 2010)

In an important respect, a pen implement that facilitates casting information in different representations supports perspective shifting in thinking about a problem, which in turn provides traction for conceptual change. Pen interfaces are a valuable tool for exploration of possible meanings through the communicative acts of sketching and writing, independent of conveying any designated meaning. As outlined in the introduction, educational interfaces that facilitate communicative actions involving representations central to a domain can maximize students' effort associated with constructing and automating new schemas (Sweller et al. 1998). It is the fertile interplay between pen input as a linguistic carrier of diverse mental representations and stimulation of cognitive flow during thinking and problem solving that make it such a promising direction for future educational interfaces.

How Can Pen Interfaces be Optimized Further for Educational Purposes?

To be developed for educational applications, digital paper and pen interfaces will require improved interface design in confirmation feedback so students have more confidence in their reliability. On the one hand, the digital pen's "quiet" low-feedback interface optimizes students' focus of attention and problem solving. On the other hand, students disliked its lack of confirmation during erasure, which caused them to question its error-proneness. In the future, digital pen interface design could benefit from confirmation feedback that is:

- Capable of minimizing cognitive load during high-load phases of problem solving, during more difficult tasks, and during mobile use when students are dual-tasking such that their base-rate of cognitive load is elevated
- Multimodal (e.g., visual, tactile, auditory), such that feedback can be adapted appropriately to a student user or usage context, and also delivered in a reinforced manner when necessary
- Infrequent, but strategically delivered only for essential needs or when solicited by a user
- Located strategically during phases of problem solving to preserve students' focus of attention

This first point refers to the fact that digital pens are expected to be used heavily in field environments and while mobile during informal learning opportunities. For this reason, sophisticated context-sensitive confirmation capabilities will be needed to support dual tasking, while at the same time preserving students' focus of attention on their field learning tasks.

Recent research has been actively exploring different types of feedback for digital pen interfaces, including tactile/vibratory, LCD visual displays embedded in the side of the pen (Livescribe 2009), miniature projection devices mounted on pens to create virtual displays overlaid on paper (Song et al. 2009), speech and non-speech auditory feedback (Leapfrog 2009), and multimodal combinations of feedback (Liao et al. 2006). The major challenge will be to develop these design directions for new digital pen interfaces, without simultaneously sacrificing students' ability to focus effectively when substantial mental effort is required for educational tasks.

Conclusions and Future Directions

In many respects, a good educational interface is like a glider that invites flight as a flow of movement while being nearly weightless and transparent, so students can experience the force of the wind without impedance. Emerging pen technologies are a promising direction for stimulating students' active self-structuring and mastery of science problem solving. They facilitate more fluent communication of non-linguistic representational content than existing keyboard-dominant interfaces, and also a parallel increase in students' ideational fluency in science (Oviatt et al. 2010). In the present research, both low- and high-performing students' ability to solve science problems was best facilitated by the digital paper and pen interface that most closely mimicked existing work practice, compared with the pen and graphical tablet interfaces. In addition, the digital paper interface most effectively facilitated skill acquisition in low-performing students. Paradoxically, all students nonetheless believed that the tablet interfaces provided best support for their performance, which indicated limited self-awareness of how to use computational tools to best support performance. In the future, it will be important to explore the potential educational benefits of digital pen interfaces in situated classroom ecosystems, over longer time periods, and for a variety of underserved and habitually lower-performing minority, indigenous, and disabled students. A concerted effort also will be required to develop new digital pen interfaces with adequate feedback explicitly for educational purposes.

Acknowledgments The second author was supported by an Incaa Designs' Science & Technology internship. Thanks to Ariana Mann, Russell Transue, Ali Maier, and Kejun Xu for assistance with data collection and analysis, and to Alex Arthur for programming support. Special thanks also to science teacher Ms. Korrie Beemer for expert assistance with problem sets and scoring protocols. We would like to thank ACM for permission to reprint Figs. 1 and 3 and Table 1.

References

Adapx (2009) www.adapx.com, Sept. 1, 2009

Aleven V, Koedinger KR (2000) limitations of student control: do students know when they need help? Proceedings of the 5th International Conference on Intelligent Tutoring Systems. Springer, London, 1839, 292–303

Anoto (2009) http://www.anoto.com/, May 1, 2009

Baddeley A (1986) Working memory. Oxford University Press, New York

Bloom P, Peterson M, Nadel L, Garrett M (eds) (1996) Language and space. MIT Press, Cambridge

Branford JD, Donovan SM (2005) How students learn: history, mathematics, and science in the classroom. National Academies Press, Washington

Bunt A, Terry M, Lank E (2009) Friend or foe? examining cas use in mathematics research. Proceedings of the 27th International Conference on Human Factors in Computing Systems. ACM Press, New York, 229–238

Cohen PR, McGee DR (2004) Tangible multimodal interfaces for safety critical applications. Special Issue Multimodal Interact, Commun ACM 47(1):41–46

Comblain A (1994) Working memory in Down's Syndrome: training the rehearsal strategy. Down's Syndr: Res Pract 2(3):123–126

Dede C (2009) Immersive interfaces for engagement and learning. Science 323:66–69

Gibson J (1977) The theory of affordances. In: Shaw R, Bransford J (eds) Perceiving, acting and knowing 3. Erlbaum, Hillsdale, pp 67–82

Lajoie S, Azevedo R (2006) Teaching and learning in technology-rich environments. In: Alexander PA, Winne PH (eds) Handbook of educational psychology, 2nd edn. Lawrence Erlbaum Association, Mahwah, pp 803–824

Larkin J, Simon H (1987) Why a diagram is (sometimes) worth ten thousand words. Cogn Sci 11:65–99

LaViola J, Zeleznik R (2004) MathPad2: a System for the creation and exploration of mathematical sketches. ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004), 23 (3), 432–440

LeapFrog (2009) http://www.leapfrog.com, April 30, 2009

Levinson S (2003) Space in language and cognition: explorations in cognitive diversity, language, culture and cognition. Cambridge University Press, Cambridge

Liao C, Guimbretiere F, Hinckley K (2005) PapierCraft: a system of interactive paper. Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology. ACM Press, New York, 241–244

Liao C, Guimbretière F, Löckenhoff C (2006) Pen-top feedback for paper-based interfaces. Proceedings of the 19th Annual ACM Symposium on User Interface Software Technology. ACM Press, New York, 201–210

Liao C, Guimbretiere F, Hinckley K, Hollan J (2008) Papiercraft: A gesture-based command system for interactive paper. ACM Transactions on Computer Human Interaction. ACM Press, New York, 14 (4), 1–27

Linn M, Lee H, Tinker R, Husic F, Chiu J (2006) Inquiry learning: teaching and assessing knowledge integration in science. Science 313:1049–1050

Livescribe (2009) http://www.livescribe.com, May 1, 2009

Luria A (1961) The role of speech in the regulation of normal and abnormal behavior. Pergamon Press, Liveright

Mestre JP (ed) (2005) Transfer of learning from a modern multidisciplinary perspective. Information Age Publishing, Greenwich

Oviatt SL (2006) Human-centered design meets cognitive load theory: Designing interfaces that help people think, Proceedings of the Conference on ACM Multimedia '06, special session on "Human-Centered Multimedia Systems". ACM, New York, 871–880

Oviatt S, Cohen A (2010) Supporting students' thinking marks: designing accessible interfaces for science education. Am Assoc Educ Res

Oviatt SL, Arthur A, Cohen J (2006) Quiet interfaces that help students think. Proceedings of the 19th ACM Symposium on

User Interface Software Technology, CHI Letters. ACM Press, New York, 191–200

Oviatt SL, Arthur A, Brock Y, Cohen J (2007) Expressive pen-based interfaces for math education. Proceedings of the Conference on Computer Supported Collaborative Learning, International Society of the Learning Sciences

Oviatt S, Cohen A, Mann A (2010) Designing educational interfaces that stimulate ideational super-fluency in science, Communications of the ACM

Paivio A (1986) Mental representations: a dual coding approach. Oxford University Press, Oxford

Pea RD, Maldonado H (2006) WILD for learning: interacting through new computing devices anytime, anywhere. In: Sawyer RK (ed) Cambridge university handbook of the learning sciences. Cambridge University Press, New York, pp 427–443

Roth WM (2005) Talking science: language and learning in science classrooms. Rowman & Littlefield, Toronto

Schwartz D, Heiser J (2006) Spatial representations and imagery in learning. In: Sawyer RK (ed) Cambridge university handbook of the learning sciences. Cambridge University Press, New York, pp 283–298

Schwartz D, Varma S, Martin L (2008) Dynamic transfer and innovation. In: Vosniadou S (ed) International handbook of research on conceptual change. Taylor & Francis, Mahwah

Shapley K, Maloney C, Caranikas-Walker F, Sheehan D (2008) Evaluation of the Texas technology immersion pilot: outcomes for the third year (2006–2007). Texas Center for Educational Research, Austin

Shapley K, Sheehan D, Maloney C, Caranikas-Walker F (2009) Evaluation of the Texas technology immersion pilot: final outcomes for a four-year study (2004–2005 to 2007–2008). Texas Center for Educational Research, Austin

Signer B (2006) Fundamental Concepts for Interactive Paper and Cross-Media Information Spaces (Doctoral dissertation, Swiss Federal Institute of Technology, 2006). Dissertation Abstracts International: C, 67(04), 1121

Song H, Grossman T, Fitzmaurice G, Guimretiere F, Khan A, Attar R, Kurtenbach G (2009) Penlight: combining a mobile projector and a digital pen for dynamic visual display. Proceedings of the 27th International Conference on Human Factors in Computing Systems. ACM Press, New York, 143–152

Sweller J, van Merrienboer J, Paas F (1998) Cognitive architecture and instructional design. Educ Psychol Rev 10:251–296

Tabard A, Mackay W, Eastmond E (2008) From individual to collaborative: the evolution of prism, a hybrid laboratory notebook. Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work, CSCW'08. ACM Press, New York, 569–578

Tsandilas T, Letondal C, Mackay W (2009) Musink: composing music through augmented drawing. Proceedings of the 27th International Conference on Human Factors in Computing Systems. ACM Press, New York, 819–828

van Merrienboer J, Sweller J (2005) Cognitive load theory and complex learning: recent developments and future directions. Educ Psychol Rev 17(2):147–177

Vygotsky L (1962) Thought and Language (MIT Press, Cambridge Ma.; Transl. by E. Hanfmann, G. Vakar from 1934 original)

Wieman CE, Adams WK, Perkins KK (2008) PhET: simulations that enhance learning. Science 322:682–683

Winne PH, Perry NE (2000) Measuring self-regulated learning. In: Boekaerts M, Pintrich P, Zeidner M (eds) Handbook of self-regulation. Academic Press, Orlando, pp 531–566

Yeh R, Liao C, Klemmer S, Guimbretiere F, Lee B, Kakaradov B, Stamberger J, Paepcke A (2006) ButterflyNet: a mobile capture and access system for field biology research. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI'06. ACM Press, New York, 571–580

Zhang H, Linn M (2008) Using drawings to support learning from dynamic visualizations. Annual Meeting of the American Educational Research Association. New York

Zucker A, Light D (2009) Laptop programs for students. Science 323:82–85