Expressive Pen-Based Interfaces for Math Education

Sharon Oviatt, Incaa Designs, 821 2nd Ave. Suite 1150, Seattle, WA 98104, USA, oviatt@incaadesigns.org
Alex Arthur, Adapx, 821 2nd Ave. Suite 1150, Seattle, WA 98104, USA, alex.arthur@adapx.com
Yaro Brock, Technical Communication Dept., Univ. of Washington, 14 Loew Hall, Seattle, WA 98195, USA
Julia Cohen, Dartmouth College, Hanover, New Hampshire 03755, USA, julia.o.cohen@alum.dartmouth.org

Abstract: Mathematics students almost exclusively use pencil and paper—that is, they learn without computational support. In this research, 16 high school students varying in ability from low to high participated in a comparative assessment of geometry problem solving using: (1) pencil and paper, (2) an Anoto-based digital stylus and paper interface, (3) a pen tablet interface, and (4) a graphical tablet interface. Cognitive Load Theory correctly predicted that as interfaces departed more from familiar work practice, students experienced greater cognitive load and corresponding reductions in their expressive fluency and planning. The results of this study indicate that students' communication patterns and meta-cognitive control can be enhanced by pen-based interfaces during math problem solving activities. In addition, low-performing students do not automatically reap the same advantage as high performers when new interface tools are introduced, which means intervention may be required to avoid expanding the achievement gap between groups unless intervention is undertaken.

Introduction

Although current graphical interfaces can support routine tasks like word processing and e-mail, they frequently fail to support more complex problem solving tasks in domains such as mathematics. In fact, current work practice for mathematics education almost exclusively involves pencil and paper, or learning without computational support. One reason for this is that modern interfaces do not support user input fluency in different representational systems (e.g., linguistic, numeric, symbolic, and diagrammatic), or flexible translation among them (e.g., from word problems and diagrams to algebraic formulas). Whereas graphical interfaces provide good support for linguistic and numeric content, symbolic and diagrammatic input are poorly supported—or not supported at all. A second reason is that traditional graphical interfaces are heavily laden with potentially distracting features. Thirdly, they typically depart from existing work practice. In the present paper, we focus on evaluating alternative interfaces that transparently mimic students' existing work practice, such that cognitive load is minimized during complex geometry problem solving tasks. We also aim to develop educational interfaces that avoid exacerbating pre-existing performance differences between low- and high-performing students, since low performers do not always benefit equally from the introduction of new computational tools due to weaker meta-cognitive skills (Oviatt, Arthur, & Cohen, 2006).

Cognitive Load Theory

Cognitive Load Theory (CLT) provides a potentially coherent and powerful basis for predicting students' performance when using new educational interfaces, and for designing educational interfaces that effectively minimize cognitive load (Mousavi, Low, & Sweller, 1995; Oviatt, 2006; Paas, Tuovinen, Tabbers, & Van Gerven, 2003; van Merriënboer & Sweller, 2005). Cognitive load involves the mental resources that a person has available for solving problems at a given time. Current work on cognitive load emphasizes limited attention and working memory capacity as specific bottlenecks that continually exert pressure on performance during information processing. Cognitive load theorists have maintained that during the learning process, students can more easily acquire new schemas and automate them if instructional methods minimize demands on their working memory, thereby reducing cognitive load (Baddeley, 1986; Mousavi et al., 1995; Paas et al., 2003; van Merriënboer & Sweller, 2005). To achieve this goal, advocates of this theory assess the "extraneous complexity" associated with instructional methods or interfaces separately from the "intrinsic complexity" associated with a student's main learning task, and then compare performance across different interfaces.

In related educational research, a multimodal presentation format has been shown to support expansion of working memory and better problem solving on geometry tasks than a single visual mode (Mousavi et al., 1995). The advantages of a multimodal presentation format for students' tutorial performance have been replicated for different tasks, dependent measures, and presentation materials, including computer-based multimedia animations (Mayer & Moreno, 1998; Tindall-Ford, Chandler, & Sweller, 1997). When using computer interfaces, it also is known that as cognitive load increases with task difficulty, users spontaneously shift to interacting more multimodally, so a flexible multimodal interface can assist users in self-managing their cognitive load (Oviatt, Coulston, & Lunsford, 2004). Furthermore, researchers have documented that performance by the same person completing the same task improves when using a multimodal interface, compared with a unimodal one (Oviatt, 1997).

In research with elementary school children and adults, active manual gesturing also was demonstrated to reduce cognitive load and improve memory during a task requiring explanation of math solutions. Furthermore, during more difficult tasks, gesturing was especially effective at minimizing cognitive load and improving memory (Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001). The physical activity of manual or pen-based gesturing is believed to play an important role in organizing and facilitating spatial information processing, thereby reducing cognitive load on tasks involving geometry, maps, etc. (Alibali, Kita, & Young, 2000; Oviatt, 1997; Rauscher, Krauss, & Chen, 1996). For an overview of other user-centered interface design techniques known to minimize cognitive load, see (Oviatt, 2006).

Adaptive Learning and Meta-Cognition

Cognitive Load Theory recently has been applied to the design of educational systems that select problem solving content of the appropriate difficulty level for a given student. In this case, the goal is to deliver an optimal level of difficulty for the student's primary learning task, rather than minimizing extraneous load associated with managing the system interface per se. This interest in student-centered tailoring of problem difficulty was inspired in part by the discovery of the expertise reversal effect, which revealed that the optimal instructional design for novices can be ineffective for more knowledgeable learners because processing redundant information overloads their working memory capacity (Kalyuga, 2006; Kalyuga, Ayres, Chandler, & Sweller, 2003; Salden, Paas, Broers, & van Merriënboer, 2004; Salden, Paas, & van Merriënboer, 2006). Although a novice may require detailed worked examples to establish new schemas, this same information actually hinders the performance of an expert for whom knowledge is already well integrated in long-term memory.

To optimize tailoring of the problem difficulty presented to a learner, some educational studies have assessed both student performance (i.e., problem correctness) and mental effort (i.e., reported subjectively) to gauge the efficiency of learning (Salden et al., 2004; van Merriënboer & Sweller, 2005). Basically, if a student solves a problem correctly and reports low effort, then their next delivered task would be more difficult. On the other hand, poor performance coupled with high effort would result in an easier problem. This research confirmed that adaptive learning protocols result in better learning progress than traditional methods, although basing adaptation on mental efficiency rather than performance alone has not shown demonstrable advantages (Salden et al., 2004). Recent research has attempted to more objectively assess a student's domain expertise as they solve problems, by evaluating the granularity of solution steps, including the number of skipped steps due to having learned and stored a schema in chunked form (Kalyuga, 2006). This pragmatic approach to calibrating a student's level of expertise as they work provides a potential future basis for real-time tailoring of educational systems.

As a related issue to expertise, research has documented that lower-performing students lack the meta-cognitive skills needed to organize and improve their own performance (Aleven & Koedinger, 2000; Winne & Perry, 2000). Among other things, such self-regulatory skills include knowing what type of problem one is working on, its difficulty level, and what type of tools or strategies are needed to solve a problem. Stronger meta-cognitive skills help students identify the best times to use computational tools, and how to use them most effectively. Past work on self-directed help systems has indicated that students frequently do not have the skills needed to utilize such resources effectively (Aleven & Koedinger, 2000). Other recent work has shown that lower-performing students are less aware than high performers of which interface tools will advance their performance best, and in some cases they prefer interface options that are *least supportive* of their performance (Oviatt et al., 2006). Given low performers' lack of savvy regarding computational tools, the introduction of new technology into classrooms risks exacerbating the existing achievement gap between low and high performers, especially if performance differences are not monitored carefully.

Pen-Based Interfaces

Pen-based interfaces have many attractive features for the education sector, including their compatibility with mobility, expressive range, suitability for collaboration, and ability to "bridge" formal, informal, and mobile learning contexts (Cohen & McGee, 2004; Leapfrog, 2006; Pea & Maldonado, 2006). Anoto-based digital stylus and paper interfaces, which span the physical and digital worlds, also are considered a promising interface for knowl-

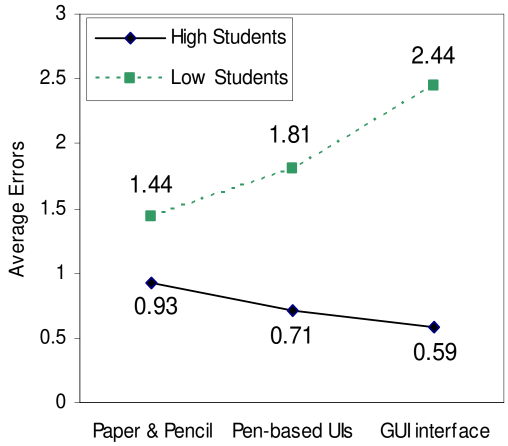

Figure 1. Difference between low- and high-performing students in math errors in the pen (DP, PT) versus graphical (GT) interfaces.

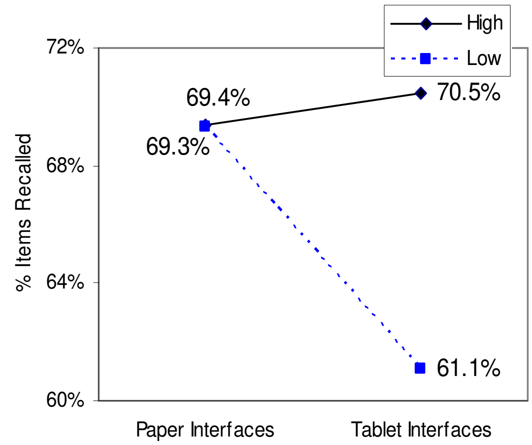

Figure 2. Percentage items recalled correctly by low- and high-performing students using paper (PP, DP) versus tablet (PT, GT) interfaces.

edge-gathering tasks in which users combine, cross-reference, and personalize information from different sources with pen-based annotations (Liao, Guimbretiere, & Hinckley, 2005).

Recent research comparing educational interfaces has shown that interfaces more similar to students' existing work practice also reduce extraneous cognitive load and improve performance during geometry problem solving tasks (Oviatt et al., 2006). A comparison of students' speed, attention, meta-cognitive control, correctness of solutions, and memory revealed that they performed better when using a digital stylus and paper interface (DP) than a pen tablet interface (PT), which in turn supported better performance than a graphical tablet interface (GT) (Oviatt et al., 2006). Cognitive Load Theory provided the basis for making quantitative rank order predictions about student performance with these different interfaces. Basically, the digital stylus and paper interface enhanced performance best because it most closely mimicked existing work practice by incorporating both pen input and the familiar, tangible paper medium. In comparison, the pen tablet interface included the pen but not the paper medium, and the graphical interface least resembled students' existing work practice. Within the math domain, both of the pen–based interfaces support a broad range of expressive input in different representational systems, including linguistic, numeric, symbolic, and diagrammatic. Such pen interfaces are particularly compatible with complex problem solving in domains like mathematics, which requires input fluency in all four representational systems and flexible translation among them to facilitate clarity of thought.

In the previously mentioned study (Oviatt et al., 2006), lower-performing students' ability to correctly solve math problems and remember the problem content they had just worked on were selectively disrupted when using the tablet interfaces, especially with the graphical tablet interface. As shown in Figure 1, high-performing students' errors did not change significantly when using the different interfaces. However, low-performing students' errors increased from 1.44 with pencil and paper (64% correct solutions), to 1.81 with the pen-based interfaces (55% correct), and 2.44 with the graphical tablet interface (just 39% correct). As shown in Figure 2, the study found parallel trends in students' recall of math content. After using paper-based versus tablet-based interfaces, the high-performing students correctly recalled 69.4% and 70.5% of the math content they had just worked on. The low-performing students recalled math content equally well after using paper-based interfaces (69.3%), but their recall dropped to 61.1% on the tablet interfaces, or 12%. From the viewpoint of CLT, the higher extraneous load involved with the tablet interfaces, especially the graphical one, derailed low performers' working memory resources from successfully solving and retaining information about the same problems.

Based on think-aloud protocols, this research also documented that the frequency with which students were distracted by the interface rather than focusing on their math increased a substantial 326% when using the pen tablet interface (e.g., "Oops, lasso didn't work") and 661% with the graphical tablet interface (e.g., "Darn, I mis-clicked"), compared with using paper and pencil. As students became more distracted with the tablet interfaces, their high-

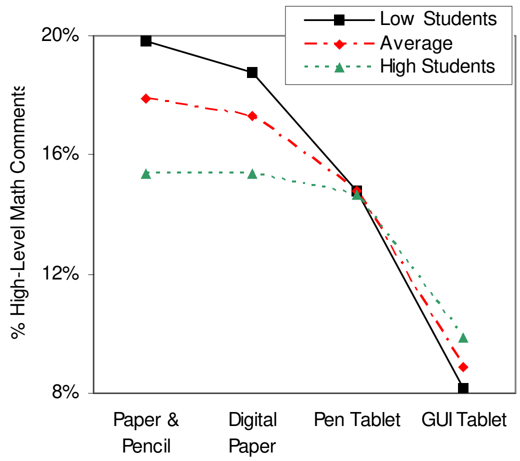

Figure 3. Percentage of high-level math comments for low- and high-performing students using different interfaces.

Table 1. Preference for the paper (PP, DP) versus tablet (PT, GT) interfaces (left), and corresponding math performance levels (right) for low- versus high-performing students.

Students	% Prefer Paper	% Prefer Tablet	% Correct Paper	% Correct Tablet
Low	37.0	63.0	57.5	50.0
High	100.0	0.0	82.5	80.0

level math comments correspondingly declined (e.g., "Oh, it's a 3D problem"), as illustrated in Figure 3. Whereas students' low-level procedural math comments were unaffected, their ability to think at a more abstract and strategic level about the nature of their math problems declined by 50.3% when they used the graphical interface, and more sharply for low-performing students (59%) than for high performers (42%). When asked which interface students would use if they had to perform their best on an AP exam, 100% of high-performing students said they would prefer the paper-based interfaces. However, Table 1 shows that for low-performing students, the reverse was true—63% said they would prefer using the tablet interfaces, even though their performance was more poorly supported by them. This performance-preference paradox reflects weaker self-regulatory skills in the lower-performing students, who clearly were less aware than high-performing students of the tools they needed to perform well (Oviatt et al., 2006).

Goals of the Study

The general goal of this study was to comparatively assess alternative interfaces with respect to their ability to minimize students' cognitive load and support successful geometry problem solving. We were specifically interested in how well different interfaces supported students' expressive fluency while thinking through solutions to problems, and any diagramming they did in advance of beginning a new problem as they clarified their understanding of what the problem meant and planned their approach to solving it. Comparisons were made for both low- and high-performing students while using: (1) existing paper and pencil work practice, (2) a digital stylus and paper interface (i.e., based on Anoto technology (Anoto Technology, 2006)), (3) a pen tablet interface, and (4) a graphical tablet interface that included a keyboard, mouse, stylus, and simplified equation editor. By collecting within-subject data on the *same students' ability to solve the same math problems,* this study aimed to provide a sensitive assessment of the relative cognitive load associated with using these alternative interfaces. Task difficulty levels varying from low to very high also were included to assess how well different interfaces supported performance across a realistic range of tasks. Both low- and high-performing students were studied so new interfaces can be designed that are accessible and supportive of learning for all students. We also were interested in examining the impact of introducing different interfaces on the performance gap between low- and high-performing students.

It was hypothesized that as interface prototypes departed more from familiar work practice, students would experience greater extrinsic cognitive load such that fewer mental reserves would be available for communicating fluently and engaging in advance planning. It also was hypothesized that higher-performing math students would experience less cognitive load than their lower-performing peers, so they would have relatively more resources available for communication and planning. In comparison with using paper and pencil, it was anticipated that introducing new interfaces also would risk magnifying the existing performance gap between high- and low-performing students, because low performers have weaker meta-cognitive skills and are less adept at using new tools.

Methods
Participants

Sixteen high school students who had recently completed a geometry class were included in the study as paid volunteers. All students used paper and pencil materials in their high school math classes, expressed an interest in technology, and were experienced users of graphical user interfaces with keyboard and mouse input. According to teacher records on students' classroom grades in geometry and also students' percentage of correct math problem solutions in this study, half of the students were classified as high-performing and half low-to-moderate. Twelve were female and four male. All students were native English speakers, although ethnic backgrounds varied.

Math Problems and Difficulty Levels

After consulting high school teachers and textbooks, math problems that students had just learned in their geometry classes were selected for the study. Teacher records of average student test performance on specific problems were used as an initial basis for classifying problems, and pilot testing then confirmed these classifications. All math problems were word problems that required translation from linguistic information into symbolic and digit-based information to solve them. Since the majority were spatially-oriented geometry problems, diagrams also were helpful in solving them. In short, successful completion of the math problems required complex problem solving using all four representational systems (linguistic, symbolic, numeric, and diagrammatic), as well as translating among them. These characteristics enabled testing the ability of different interfaces to support flexibly expressive communication patterns, which are required for extended problem solving in domains like geometry. The number, format, and type of information varied in problems of different difficulty levels, such that harder problems involved more steps to solution, information presented in different formats (e.g., integers versus ratios), incidental information not required for solution, and so forth. For further detail on problem sets, see (Oviatt et al., 2006).

Procedure

Students were tested in pairs and given instructions and practice together. They were told that their input regarding the different interfaces would be used to design a math camp for younger children. The student volunteers were shown the four different sets of materials that they would use to solve problems, including: (1) standard pencil and paper, (2) digital stylus and paper (i.e., Nokia stylus with Anoto-based paper technology), (3) tablet computer with stylus input, and (4) tablet computer with keyboard, mouse, and stylus input, which was enhanced with a simplified MathType equation editor containing 11 symbols not on the keyboard (e.g., square roots, powers).

For all four conditions, each problem set was presented on a Toshiba Portege laptop screen, as shown in Figure 4, which included the main word problem (top) along with any terms or equations required to complete the problem (bottom left). In the two paper-based conditions, students simply read the problem on the computer screen but did their work on paper. In the two tablet-based conditions, they entered their work on the computer using Windows Journal for the pen tablet condition, and either MathType or Windows Journal (i.e. using a stylus) for the mixed graphical tablet interface. Figure 4 shows the graphical tablet interface condition, with MathType (left side) and Windows Journal (right side) both open. In the pen tablet condition, Windows Journal was the only input area open, and in the two paper conditions the middle of the screen shown in Figure 4 was blank. In all conditions, students were told they could use their calculator and were free to use their materials any way they liked. With the graphical tablet interface, they could use the keyboard and equation editor or pen input however they wished.

For each of the three computer interfaces, students were given instructions on how it worked and allowed to practice until they were familiar and had no more questions. Beyond orientation, students were told to work at their own pace and concentrate on solving each problem. If they couldn't complete a problem, they were instructed to go to the next. Each student completed 16 math problems during the main test session, four problems apiece in each of the four conditions.

Research Design

This study involved a mixed factorial experimental design, with within-subject independent factors including: (1) Type of Interface: paper and pencil hardcopy materials (PP), digital stylus and paper interface (DP), pen tablet interface (PT), and graphical tablet interface (GT), and (2) Math Problem Difficulty Level: low, moderate, high, and very high. Each student completed a set of four problems per condition, which increased progressively in difficulty. The specific content of different problem sets and order of presentation of the interface conditions were counterbalanced. The main between-subject factor was: (3) Student Performance Level: high, low.

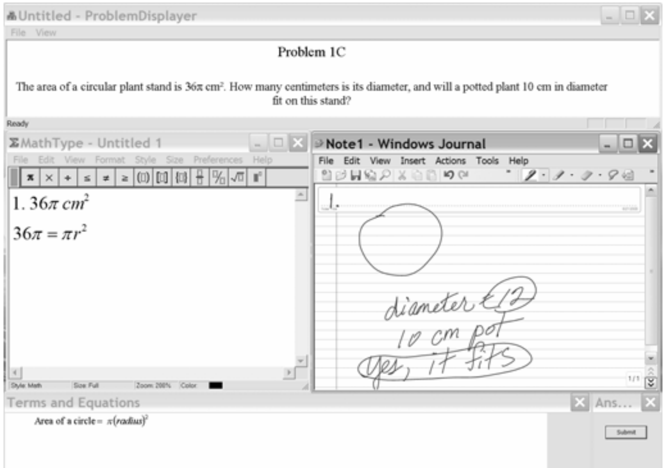

Figure 4. Interface used to display math problems

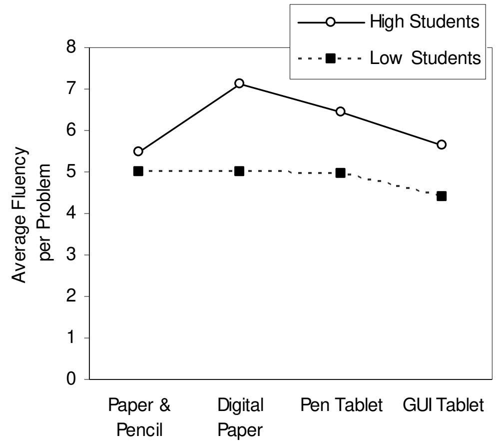

Figure 5. Fluency for high- versus low-performing students in different interfaces

Dependent Measures and Coding

Fluency in Different Representational Systems

The number of (1) words (including abbreviations), (2) digits, (3) symbols (e.g., π), and (4) diagrams that students generated while working on each problem was totaled and then summarized as an average number per problem in each condition.

Advance Planning Prior to Problem Solving

In the domain of geometry, diagramming of the spatial relations among objects is a common initial step that helps students to clarify their understanding of the problem and prepare to work. The number of diagrams that each student produced was totaled and summarized as an average number per problem.

Reliability

Fluency counts were scored again by two independent coders for 13% of the data. These counts matched exactly 93%, 97%, 94%, and 100% of the time for linguistic, numeric, symbolic, and diagram counts, respectively.

Results

Data were available on 256 problem solutions for the dependent measures reported below.

Fluency Using Different Interfaces

For high-performing students, their expressive fluency while solving math problems using different interfaces increased from an average of 5.47 per problem using pencil and paper, to 7.13 in the digital stylus interface and 6.43 with the pen tablet, but dropping back to 5.65 when using the graphical tablet interface. For low-performing students, fluency remained more stable at 5.02, 5.02, 4.95, and 4.42, respectively, for the same interfaces. Figure 5 illustrates that the high-performing students were significantly more fluent when using the two pen-based interfaces (mean = 6.78) than with the other interfaces (mean = 5.56), paired t test, $t = 2.06$ ($df = 7$), $p < .04$, one-tailed, or 22% more fluent. The main source of increased fluency when they used the pen-based interfaces involved producing 2.3 more symbols and 1.5 more digits per problem. In contrast, fluency of the low-performing students did not change when they used the pen-based interfaces compared with the others, paired $t < 1$. Fluency levels also did not differ significantly between paper and pencil and the graphical tablet interface, paired t test, $t < 1$.

The high-performing students' average fluency was 6.17 per problem, compared with 4.85 for the low-performing students, a marginal difference between groups across all interfaces, independent t test, $t = 1.75$ ($df = 14$), $p < .051$, one-tailed. The high-performing students were significantly more fluent than the low performers when using the digital paper and stylus interface, independent $t = 1.99$ ($df = 14$), $p < .035$, one-tailed, and also when using the pen tablet interface, independent $t = 1.93$ ($df = 14$), $p < .04$, one-tailed. In fact, the high performers averaged 36% more expressive fluency when using the pen-based interfaces, compared with the low-performing students. In comparison, these groups did not differ significantly in fluency when using pencil and paper, independent $t < 1$, the

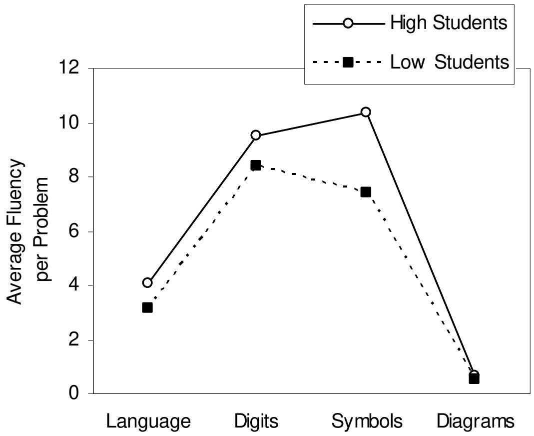

Figure 6. Average fluency for high- versus low-performing students using different types of representational system.

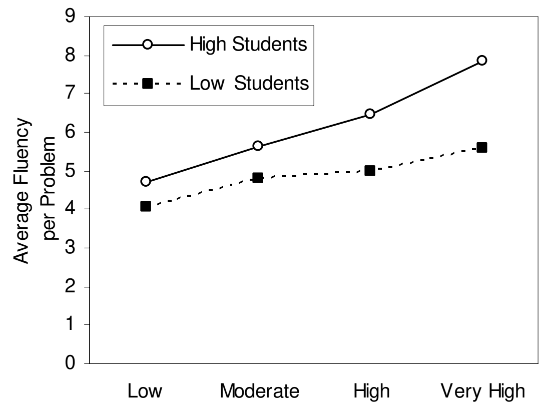

Figure 7. Average fluency for high- versus low-performing students due to task difficulty.

graphical interface, independent t = 1.12, N.S., or even pen input within the graphical interface, independent t < 1. When using the graphical tablet interface in which students had free choice to type or use pen input, 50% of students mixed text and pen input (e.g., using text for digits and formulas, pen for diagramming and labeling), 37.5% provided only pen input, and 12.5% only used text. Overall, 37% of all input was text and 63% pen input.

Fluency in Different Representational Systems and Task Difficulty Levels

As shown in Figure 6, students actively used all four representational systems while solving geometry problems. High performers averaged 4.08 linguistic, 9.54 numeric, 10.38 symbolic, and .67 diagrammatic content per problem, while low performers averaged 3.16 linguistic, 8.37 numeric, 7.38 symbolic, and .51 diagrammatic content. The high performers were significantly more fluent than low performers when using the challenging symbolic content (means 10.38 and 7.38, respectively), independent t = 2.02 (df = 14), p < .035, one-tailed, a 41% increase. However, the groups did not differ in other fluency rates.

As problems became more difficult, high-performing students' fluency increased steadily from 4.73 on low difficulty problems, to 5.63 on moderate, 6.46 on high, and 7.86 on very high difficulty ones. Likewise, for low-performing students, fluency increased from 4.05 on low, 4.78 on moderate, 4.97 on high, and 5.60 on very high difficulty problems. These shifts in fluency represented a significant increase between low and moderately difficult problems, paired t = 3.35 (df = 15), p < .002, one-tailed, moderate and high difficulty problems, paired t = 2.31 (df = 15), p < .02, one-tailed, and high and very high difficulty, paired t = 2.09 (df = 15), p < .03, one-tailed. Compared with low and moderate difficulty problems, on the high and very high difficulty ones students' diagramming increased by 126%, digits by 55% and symbols 27%, whereas linguistic content actually declined 5%.

As illustrated in Figure 7, the high- versus low-performing students also diverged more in their fluency as problem difficulty increased, which became most apparent on the high and very high difficulty problems. While the groups did not differ in fluency at the low and moderate difficulty levels (t < 1 and t = 1.34 N.S.), the high-performing students were marginally more fluent than low performers at the high difficulty level (independent t = 1.65 (df = 14), p < .065, one-tailed), and they were significantly more fluent than low performers at the very high difficulty level (t = 1.95 (df =14), p < .04, one-tailed). As problem difficulty increased from low to very high, the high-performing students increased their fluency by 66%, while low performers only increased by 38%. On the very high difficulty problems, high-performing students were 40% more fluent, on average, than the lower performers.

Planning Prior to Problem Solution

Ninety-four percent of students engaged in diagramming before or during their math problem solutions. Low- and high-performing students exhibited no significant difference in frequency of diagramming, independent t = 1.04 (df = 14), N.S., but considerable individual differences were evident among students. As shown in Figure 8,

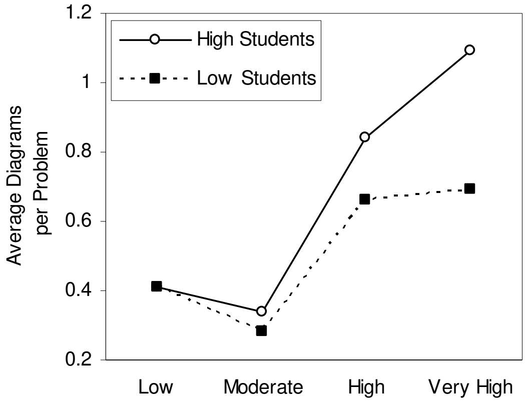

Figure 8. Average number of diagrams per problem as a function of task difficulty.

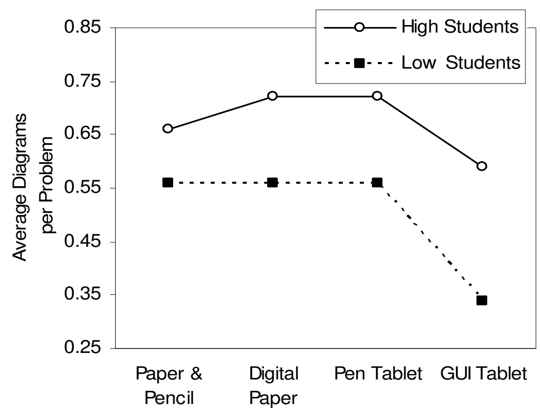

Figure 9. Average number of diagrams per problem as a function of interface.

the average number of diagrams increased with task difficulty for both low performers (means = .41, .28, .66, and .69 for low to very high) and high performers (means = .41, .34, .84, 1.09). A paired t test confirmed that diagramming increased significantly between low/moderate and high/very high difficulty problems, $t = 6.30$ (df =15), $p < .001$, one-tailed. Separate analyses also indicated that both high- and low-performing students significantly increased their diagramming on the high/very high difficulty problems, $t = 5.06$ (df = 7), $p < .0005$, one-tailed, and $t = 4.08$ (df = 7), $p < .0025$, one-tailed, respectively. It is noteworthy that high performers increased their diagramming 158% on the harder math problems, whereas low performers only increased 95%. In addition, a linear regression between task difficulty and the likelihood of diagramming revealed a correlation of .90, with 82% of the variance in students' likelihood of diagramming accounted for by knowing the difficulty level of their math problem.

With respect to diagramming in different interfaces, Figure 9 illustrates that high-performing students averaged .66 diagrams per problem when using pencil and paper, .72 with digital stylus and paper, and .72 with the pen tablet, but dropped to .59 with the graphical tablet—none of which were significant differences, ts< 1. Low performers remained stable at .56 diagrams in all interfaces except the graphical one, for which they dropped significantly to .34, Wilcoxon Signed Ranks test, $z = 1.75$, $p < .04$ (one-tailed), a 39% drop.

Discussion

As shown in Figure 6, students actively used all four representational systems while solving geometry problems. This highlights the importance of developing more powerfully expressive pen interfaces for supporting educational domains like math, which require symbolic and diagrammatic input as well as linguistic and numeric. In addition, 94% of students drew diagrams before solving their problems, and they increased diagramming 117% between low and very high difficulty problems. The pen interfaces both supported diagramming at levels as high as existing pencil and paper work practice, although diagramming dropped 22% when students used the graphical tablet interface—in fact, more sharply by 39% for the low-performing students. Although students in this study were all expert graphical interface users, and the mixed graphical tablet interface also supported pen input, they still used this interface less fluently and with less foresight than the two pen interfaces. This finding is consistent with previous research revealing weaker meta-cognitive skills in low performers (Winne & Perry, 2000), and also less high-level planning among low performers when using a graphical tablet interface (Oviatt, 2006).

As predicted by Cognitive Load Theory, high performers experienced less cognitive load than lower performers when working on the same math problems. As such, they had more mental resources available for increasing their fluency level appropriately as interfaces and problems increased in difficulty. Compared with low-performing students, they were 40% more fluent on the very high difficulty problems, and 41% more fluent with symbolic content. In addition, the higher performers actually were *super-fluent* when using the two pen interfaces—the digital paper and pen interface, and pen tablet interface. They became 36% more fluent with these pen interface tools, although the low performers were not similarly stimulated. This difference between groups in their use of the pen interfaces is important because the activity of self-expression itself can serve to clarify thought.

One objective of geometry teachers is to encourage students to diagram more frequently to facilitate their problem solutions. Like expressive fluency, diagramming can function as a self-organizing activity that assists students in planning clearer problem solutions. Typical student comments about diagramming included: "I'm a visual learner. I like to draw pictures to help me think clearly." And "I need visualizations to figure out the problems." While diagramming is particularly well supported by the more expressively powerful pen interfaces, higher-performing students were more likely to take full advantage of this capability. The high performers specifically responded to harder math problems by diagramming 158% more than on easier problems, compared with just a 95% increase for low performers. This indicates that low performers may need instruction to encourage higher levels of diagramming as an aid to solving difficult problems, and to ensure that they make full use of the pen interfaces.

Table 2.

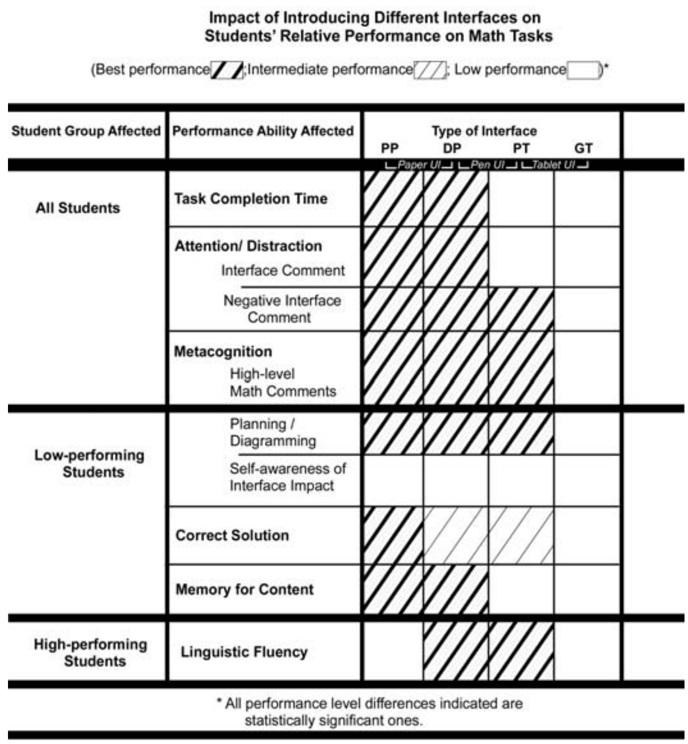

Table 2 summarizes the convergent pattern of results that has emerged based on the present and previous studies that examined the impact of different interfaces on students' geometry performance (see (Oviatt et al., 2006) for discussion of previous findings). Analyses from both studies consistently reveal that meta-cognitive behavior (i.e., diagramming, high-level math comments) decline when using the graphical tablet interface, with advance diagramming specifically reduced in the low-performing students. The present study also showed that high-performing students were super-fluent when using pen interface tools, although low-performing students did not realize the same advantage of these interfaces. As shown in Table 2, the convergent results that emerge from the present and previous studies indicate that the paper and pen interface (DP) supported performance the best of all interfaces compared, with no overall disadvantages compared with paper and pencil work practice. As such, it provides the most viable interface option for introducing digital tools into complex math problem solving activities. The pen tablet interface (PT) was the next most effective, and the graphical tablet interface (GT) least effective. These interface differences are reflected in decreasing advantages from the left to right side of Table 2.

During educational activities, students work on learning to master tasks that stretch existing capabilities and create a relatively high baseline level of cognitive load. For this reason, educational tasks present an ideal forcing function for developing interfaces that minimize load. In the field of math education, it will be especially important for educators to participate in developing new interfaces, especially for weaker students, to ensure that new technologies are developed that do not exacerbate pre-existing performance differences between groups.

References

Aleven, V., & Koedinger, K. R. (2000). Limitations of student control: Do students know when they need help? In *Proceedings of ITS 2000*. 292-303.

Alibali, M., Kita, S., & Young, A. (2000). Gesture and the process of speech production: We think, therefore we gesture. *Language and Cognitive Processing, 15*, 593-613.

Anoto Technology. (2006). Retrieved Feb. 23, 2007: http://www.anoto.com

Baddeley, A. (1986). *Working Memory*. New York: Oxford University Press.

Cohen, P. R., & McGee, D. R. (2004). Tangible multimodal interfaces for safety critical applications. *Special Issue on Multimodal Interaction, Communications of the ACM, 47*(1), 41-46.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. J., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science, 12*(6), 516-522.

Kalyuga, S. (2006). Rapid cognitive assessment of learners' knowledge structures. *Learning and Instruction, 16*, 1-11.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 36*(1), 23-31.

Leapfrog. (2006). Retrieved Nov. 8, 2006: http://www.leapfrog.com

Liao, C., Guimbretiere, F., & Hinckley, K. (2005). PapierCraft: A system of interactive paper. In *Proceedings of UIST'05*, Seattle, WA. ACM Press: 241-244.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*(2), 312-320.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319-334.

Oviatt, S. L. (1997). Multimodal interactive maps: Designing for human performance. *Human Computer Interaction, 12*(1-2), 93-129.

Oviatt, S. L. (2006). Human-centered design meets cognitive load theory: Designing interfaces that help people think. In *Proceedings of ACM Multimedia '06, special session on "Human-Centered Multimedia Systems"*. Santa Barbara, CA. ACM Press: 871-880.

Oviatt, S. L., Arthur, A. M., & Cohen, J. (2006). Quiet interfaces that help students think. In *Proceedings of UIST '06*, Montreaux, Switzerland. ACM: 191-200.

Oviatt, S. L., Coulston, R., & Lunsford, R. (2004). When do we interact multimodally? Cognitive load and multimodal communication patterns. In *Proceedings of ICMI'04*, Penn State University, PA. ACM: 129-136.

Paas, F., Tuovinen, J., Tabbers, H., & Van Gerven, P. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*(1), 63-71.

Pea, R. D., & Maldonado, H. (2006). WILD for learning: Interacting through new computing devices anytime, anywhere. In R. K. Sawyer (Ed.), *Cambridge University Handbook of the Learning Sciences* (pp. 427-443). New York: Cambridge University Press.

Rauscher, F., Krauss, R., & Chen, Y. (1996). Gesture, speech and lexical access: The role of lexical movements in speech production. *Psychological Science, 7*, 226-231.

Salden, R. J. C. M., Paas, F., Broers, N. J., & van Merriënboer, J. J. G. (2004). Mental effort and performance as determinants for the dynamic selection of learning tasks in air traffic control training. *Instructional Science, 32*, 153-172.

Salden, R. J. C. M., Paas, F., & van Merriënboer, J. J. G. (2006). A comparison of approaches to learning task selection in the training of complex skills. *Computers in Human Behavior, 22*, 321-333.

Tindall-Ford, S., Chandler, P., & Sweller, J. (1997). When two sensory modes are better than one. *Journal of Experimental Psychology: Applied, 3*(3), 257-287.

van Merriënboer, J., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review, 17*(2), 147-177.

Winne, P. H., & Perry, N. E. (2000). Measuring self-regulated learning. In M. Boekaerts, P. Pintrich & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 531-566). Orlando, Florida: Academic Press.

Acknowledgments

Thanks to Rachel Coulston, Marisa Flecha-Garcia, and Rebecca Lunsford for recruiting and scheduling students, and composing and pilot testing math problems. This research was supported by DARPA Contract No. NBCHD030010. Any opinions, findings, or conclusions are those of the authors, and do not necessarily reflect the views of DARPA or the Department of the Interior.