THE 1984 OLYMPIC MESSAGE SYSTEM: A TEST OF BEHAVIORAL PRINCIPLES OF SYSTEM DESIGN

There was more than athletic talent being pressed to peak performance at the 1984 Olympic Games in Los Angeles. Behind the scenes, a multilingual Olympic Message System ran round-the-clock keeping more than 10,000 athletes and officials in contact with families and friends, both far and near.

JOHN D. GOULD, STEPHEN J. BOIES, STEPHEN LEVY,
JOHN T. RICHARDS, and JIM SCHOONARD

This study is a comprehensive research effort aimed at evaluating a computer system design methodology. It reports on the 1984 Olympic Message System (OMS), a voice mail system that was developed according to three behavioral principles. It describes a project from start to finish—from design and development to actual use by the customer. This research is unique in that part of its purpose was to carry out a case study of system design methodology. Consequently, the research effort involved keeping a diary; recording observations, results, and personal feelings; retaining early versions of materials; and building a usage analysis recording system in the final product and carrying out the analyses later.

Fifteen behavioral methodologies used to achieve good usability are described. The dates, times, methodologies used, and numbers of people involved, as well as results, explain how the system development actually proceeded. All aspects of usability evolved in parallel and under one focus of responsibility. We also mention some mistakes made and how the behavioral methodologies allowed us to identify and recover from them. This makes it possible for the reader to learn when and how a particular methodology contributes to the design process, and how long it may take to carry out.

PRINCIPLES OF SYSTEM DESIGN

In the past decade, we have been trying to arrive at procedures that could be used to develop computer-based systems that are reliable, responsive, easy to learn, useful, and desirable. We have recommended three principles [10, 11] to test this research:

(1) *Early focus on users and tasks.* Designers must first understand who the users will be. This understanding is determined in part by directly studying their cognitive, behavioral, anthropometric, and attitudinal characteristics, in part by studying the nature of the expected work to be accomplished, and in part by making users part of the design team through participative design or as consultants.

(2) *Empirical measurement.* Early in the development process, intended users' reactions to printed scenarios and user manuals should be observed and measured. Later on they should actually use simulations and prototypes to carry out real work, and their performance and reactions observed, recorded, and analyzed.

(3) *Iterative design.* When problems are found in user testing, as they will be, they must be fixed. This means design must be iterative: There must be a cycle of design, test and measure, and redesign, repeated as often as necessary. Empirical measurement and iterative design are necessary because designers, no matter how good they are, cannot get it right the first few times (see [11]).

SO WHAT? FIRST IMPRESSIONS

When we describe these principles, we often hear the following reactions:

- *"They're obvious. Everybody says that."* This reaction is simply incorrect. Of 450 system designers and developers who were asked to write down the steps they recommend in the design of an office system, 26 percent of them mentioned none of the three, and another 35 percent mentioned only one of the three principles [11].

- *"Everybody does these things."* We pay close attention to talks describing the development of one system or another, and find major differences between what we recommend and what actually happens. Often designers tell us that they intended to follow these principles, but for one reason or another, they were prevented from following them. Occasionally designers think they are following these principles when they are not (see [11]). Some design approaches are *milestone oriented* (e.g., project reviews, phase reviews), some are *specification oriented* or *document oriented* (e.g., functional descriptions, chart talks), and some are *characteristics oriented* (e.g., the system should be consistent, have a desktop metaphor). Our design principles are *process oriented*; that is, they specify an empirical approach to design, regardless of the type of system.
- *"Human factors is just fine-tuning."* Clayton Lewis has characterized this view as the "peanut butter" theory of usability: You can spread it on at the end, like peanut butter. However, like lukewarm chicken soup, it is bland and not curative. Gloss does not fix design defects.
- *"In real life, you can't follow them."* We often hear this reaction from designers who insist that following them takes too long. They say that the principles do not work on big projects or that they cannot test a system before it exists. On several occasions we have heard descriptions of development projects in which the initial intention was to follow these principles, but for various practical reasons, the developers discovered they could not follow them.
- *"You can't measure usability."* This is not correct. You can measure usability by (1) having in mind at the outset some behavioral criteria that your system should meet—for example, the system should be easy to use; (2) putting these criteria into specific, testable terms; and (3) testing against these specifications as development proceeds. For example, with a point-of-sale terminal a file clerk from a temporary agency with 30 minutes of training should be able to process customers' purchases with either credit cards, checks, or cash in 1.5 minutes per customer. This should be accomplished by asking no more than an average of 0.3 questions of a supervisor per instance. These specifications then become design goals, analogous to other design goals, for example, memory swap time, installation time, expected service calls, expected cost, and selling price.

THE IMPORTANCE OF THE STUDY

Building OMS was particularly difficult because of the high risk involved and the need for extremely good usability. The Olympics were in the public eye, and OMS could be used by anyone in the world. It was very visible and had the ability to handle sensitive information (personal communications). It was potentially subject to sabotage and abuse, and could have failed in many different and public ways. OMS had to work right and well when the gates opened. The dates could not slip. There would be no second chance to improve it.

This case study demonstrated that by following these principles a large system (a network of over 35 computers scattered over the Los Angeles area) can be rapidly designed, developed, tested, and changed (eight months) by a small group of people with a successful outcome for users.

THE OLYMPIC MESSAGE SYSTEM

There were approximately 10,000 Olympic athletes and officials in Los Angeles. The athletes lived in dormitories in two villages on the large university campuses of USC and UCLA. As was true of previous Olympic games, the athletes lacked support groups in the villages. Their families, friends, and personal coaches did not live there, and many did not travel to Los Angeles. It was estimated that over 50 different languages could be heard in Los Angeles during the games.

What Was the Olympic Message System?

OMS was built upon the IBM Audio Distribution System (ADS) code base [8, 9] and ran on six connected IBM S/1 computers. It allowed Olympians (the main user group) to send and receive voice messages among themselves. They could hear a message in the sender's own voice—as soon as it was recorded and exactly as he or she said it. They could use OMS from almost any push-button telephone in the world. People from around the world could send messages to the athletes and officials. Figure 1 (next page) is an example of a parent in Ireland leaving a message for his competing son. Over half of the messages Olympians received came from parents, friends, former coaches, etc., who could not come to Los Angeles (see [4] for a usage summary). Figure 2 (next page) is an example of an American Olympian listening to a message sent from his father. OMS had an entirely prompted user interface. Figure 3 (next page) is an example of an American Olympian leaving a message for an Olympian from Australia.

Since many people around the world who might wish to leave a message for an Olympian would not have access to a push-button telephone, OMS had to work with dial telephones as well. Non-Olympians called their own country's National Olympic Committee

OMS was built upon the IBM Audio Distribution System (ADS) code base and ran on six connected IBM S/1 computers. It allowed Olympians to send and receive voice messages among themselves.

Caller: (Dials 213-888-8888.)
Operator: Irish National Olympic Committee. Can I help you?
Caller: I want to leave a message for my son, Michael.
Operator: Is he from Ireland?
Caller: Yes.
Operator: How do you spell his name?
Caller: *K-E-L-L-Y.*
Operator: Thank you. Please hold for about 30 seconds while I connect you to the Olympic Message System.
Operator: Are you ready?
Caller: Yes.
OMS: When you have completed your message, hang up and it will be automatically sent to Michael Kelly. Begin talking when you are ready.
Caller: "Michael, your Mother and I will be hoping you win. Good luck." (Caller hangs up.)

FIGURE 1. Example of a Parent Leaving a Voice Message for an Olympian

You: (Dial 740-4560.)
OMS: Olympic Message System.
Please keypress your three-letter Olympic country code.
You: U S A
OMS: United States. Etats-Unis.
Please keypress your last name.
You: J O N E
OMS: John Jones.
Please keypress your password.
You: 4 0 5
OMS: New messages sent by Message Center.
"John, good luck in your race. Dad."
End of message.
Press 1, listen again; 2, leave a message; 3, hang up.
You: 3
OMS: Good-bye.

FIGURE 2. Example of You (a user) Listening to a Message

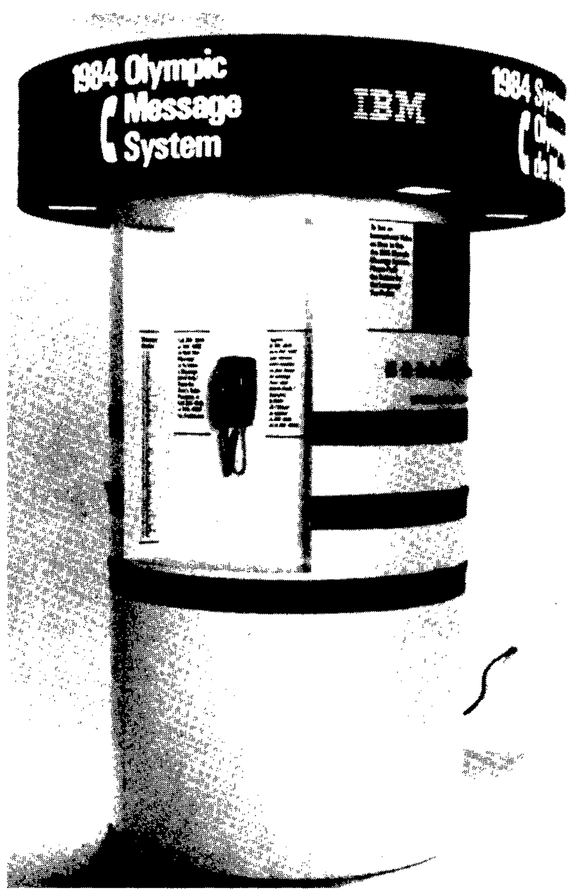
(NOC) office in Los Angeles. This ensured they would speak to someone who understood their language. A staff member, using a push-button telephone, connected the caller to OMS, aided the caller in any other way, and then got off the line. The caller spoke his or her message and hung up. The voice message was immediately in the new message box of the appropriate Olympian. If the NOC office was not staffed, the call was forwarded to a central group of telephone operators who connected the caller to OMS for that Olympian.

It was impossible to train non-Olympian callers. In order for them to become familiar with OMS, we prepared a one-page "Family and Friends User Guide" for the International Olympic Committee to send to each country's NOC. In turn, they would send them to each Olympian who could then give them to anyone he or she chose. Each pamphlet contained the telephone number of that country's NOC office in Los Angeles. As a backup, each Olympian was given postcards, printed in the appropriate language, upon arrival in Los Angeles. These postcards explained how to use OMS to send a message to an Olympian. There was also room for a personal note. After the postcard was filled out, IBM filled in the telephone number of the Olympian's own NOC office and mailed it by express mail. Olympians sent over 20,000 postcards.

OMS worked in 12 languages. To the best of our knowledge, it was the first computer system to work in several national languages simultaneously. The languages were Arabic, English, French, German, Italian, Japanese, Korean, Mandarin, Norwegian, Portuguese, Russian, and Spanish. OMS kiosks (Figure 4) were located in about 25 places in the Olympic villages, as well as at other important locations. These kiosks provided a visual presence for OMS. Each contained a PC-driven visual display of the names of Olympians who had new messages, an electronic bulletin board that displayed news items of interest, and a videodisc of a mime demonstrating how to use OMS. Olympians could press 1 of 12 buttons on the kiosk to select the appropriate language to hear the associated audio.

You: (Dial 740-4560.)
OMS: Olympic Message System.
Please keypress your three-letter Olympic country code.
You: U S A
OMS: United States. Etats-Unis.
Please keypress your last name.
You: J O N E
OMS: John Jones.
Please keypress your password.
You: 4 0 5
OMS: No new messages. Press 1, leave a message; 2, listen to an old message; 3, hang up.
You: 1
OMS: Please keypress the country code of the person you want to leave a message for.
You: A U S
OMS: Australia. Australie.
Please keypress recipient's last name.
You: B R O W
OMS: Jane Brown.
Press 1 when you have completed your message.
Begin talking when you are ready.
You: "I'll meet you tonight at 8:00."
1
OMS: Press 1, listen to this message; 2, send it; 3, do not send it.
You: 2
OMS: Message sent to Jane Brown.
Press 1, leave a message; 2, listen to an old message; 3, hang up.
You: 3
OMS: Good-bye.

FIGURE 3. Example of You (an Olympian) Sending a Message



There were about 25 OMS kiosks located around the Olympic village. The voice mail system was the first to work in 12 different languages simultaneously.

FIGURE 4. An Olympic Message System Kiosk

Kiosks also contained copies of the "Olympic Message System User Guide" printed in 12 national languages.

All signs on kiosks were in English and French—the two official languages of the Olympics. Kiosks identified a Help Line number that Olympians who were having trouble could call. If need be, the person who answered could add on OMS, as a third party, and press the keys for the Olympian, thus helping him or her, for example, to sign on. This outreach program was much broader and more integrated than most training programs.

THE SUCCESS OF OMS

In evaluating the merits of our principles of design, one must ask if OMS was successful. It was. It was reliable and responsive. It ran 24 hours per day for the four weeks the Olympic villages were open. The system was never down entirely. We worked in the villages everyday and observed that the Olympians liked the system. Also, as shown in Table I, it was used a lot. Forty percent of Olympians used it at least once. It was used an average of 1–2 times per minute, 24 hours per day (see [4] for a full description of usage, including by country and by language).

CARRYING OUT THE BEHAVIORAL PRINCIPLES OF DESIGN

We followed the three behavioral principles mentioned above. We discovered that the extra work these principles initially require greatly reduces work later on. It is not always easy, however, to follow the principles. At times it was psychologically difficult to break away from the terminal or leave the lab. There were also personal conflicts, for example, deciding whether to spend time programming a simulator or get on with the "real work" of programming the system.

How Did We Follow Them?

Printed Scenarios. At the outset, in December 1983, we prepared printed scenarios of exactly how we envisioned the user interface would look. They were similar to Figures 1–3. These scenarios contrast with the more typical approach of preparing a list of functions that a new system will contain. They provided the first definition of system function and the user interface. In ways that are hard to imagine, they powerfully determined deep system organization. They identified conflicts that a list of functions could not do. They allowed people to see, comment, and criticize at a time when their comments could have the most impact. They gave an existence to the system. They were in a form that behavioral rationale for each step could be carefully examined—by designers, prospective users, and management. The Olympic Committee was clearly impressed by this behavioral orientation in our status presentations.

The scenarios provided the opportunity to make changes to the potential user interface, and accompanying function, before any code was written. In particular, they helped to define the sign-on procedure. They drove significant system organizational considerations;

TABLE I. Summary of OMS Usage

	Totals	Means per Olympian
Number of Olympian sign-ons	31,407	6.83
Number of non-Olympian sign-ons	11,778	
Total number of usages	43,185	9.39
Messages sent from		
Same country	4,035	0.88
Other countries	1,678	0.36
Self	1,151	0.25
Non-Olympians	11,778	2.56
Total messages sent	18,642	4.05
Messages listened to	17,213	3.74

Note that the data are based on 4,601 Olympians who used OMS from around July 14 to August 13. An additional 1,648 messages were sent to Olympians who either did not use OMS or did not arrive at Los Angeles. The means are for each of these 4,601 Olympians.

In evaluating the merits of our principles of design, one must ask if OMS was successful. It was. It was reliable and responsive. It ran 24 hours per day for the four weeks the Olympic villages were open. The system was never down entirely.

for example, we dropped the message verification function whereby after sending a message a user could learn if and when it was heard by the recipient. We dropped the ability to send the same message to a distribution list of users, as well as the function of OMS calling Olympians who had new messages. These scenarios saved time, because, on the basis of people's feedback, code was never written for functions that otherwise would have been implemented. Later, when working at various depths of coding and preoccupied with details at a microlevel, these printed scenarios provided a useful high-level reminder of exactly what we were trying to do.

Early Iterative Tests of User Guides. There were two main user groups: Olympians, and family and friends. Writing user guides early provides a portable way to bring the essence of a new system to the attention of potential users in a form that they can react to. It is a useful elaboration to user scenarios. Based on feedback from various tests of people using OMS, we iterated over 200 times on the English version of the user guide for Olympians, called the "Olympic Message System User Guide," and over 50 times on the English version of the "Family and Friends User Guide." How could we do this many iterations? Requests and negotiations with others were not required. The "Family and Friends User Guide" was finalized in April so that it could be sent to the Olympic Committee in time to be mailed to the homes of Olympians. Testing and modification of the "Olympic Message System User Guide" continued throughout June.

Early versions of the two guides were also written before coding began. They served to identify issues and problems in system organization. For example, the user interface for family and friends would require trained intermediaries who would have to work very fast so as to minimize the expensive long-distance telephone charges callers would incur when calling from outside the United States. Another example was the realization that OMS operators needed to be able to change the national language that an Olympian heard OMS in if he or she wanted a different one than was selected by his or her NOC, for example, a French-speaking Canadian.

The brief user guides became the definitive OMS documents. This has happened before, for example, in the 1970s with early versions of IBM's ADS [8], at Wang with a text editor, and at Digital Equipment Corporation with the VAXstation (personal communication from R. Rubenstein, 1984; see also [13]).

An unanticipated consequence of testing and iterating is that they prevent well-intentioned but counterproductive changes later on. On more than one occasion before the Olympics started, a powerful person proposed a significant, but poor, rewrite of the "Olympic Message System User Guide." Each time, we encouraged the person to do the proposed rewrite or have one of his people do it, and then test it out on 25 or so users, comparing the results with those we had already obtained. Our proposals had significant impact. There was more appreciation of and respect for the empirical steps that had led to the existing version.

Early Simulations. A few weeks after the project started, we ran simulations of the English version of OMS on an IBM VM system, using IBM 3277 and 3279 terminals. The PF keys on these terminals are laid out in the same 3-by-4 arrangement as push-button telephone keys. Participants pressed these instead of telephone keys to give commands. They would speak their brief messages, and the experimenter would type them. OMS prompts and feedback were displayed on the CRT screen. The experimenter would read the prompts and messages aloud, so participants would hear (rather than see) them.

This was more involved than a typical simulation. The program written on VM formed a real OMS, but a displayed version rather than an audio version. The code could be transferred in a matter of minutes, exactly "as is," to the harder-to-program-and-debug IBM S/1 computer. There it ran with no modification and provided the audio version of OMS. This VM approach was originally developed for ADS (see [8] and [9]). It was possible to use quickly because the table-driven ADS interface was easy to modify, and we were able to use much of the ADS code base in the simulator. This Voice Toolkit (see [12]) allowed us to debug the user interface, conduct informal user experiments for the interfaces for both major user groups (Olympians, and family/friends), and provide demonstrations to elicit people's comments.

We aimed at getting novices characteristic of both user groups to be able to read a one-page description of OMS and then carry out simple test problems. As usual, the approach was to test OMS with experimental participants, modify it, and test it again. In the early stages, laboratory personnel and visitors were sufficient to get rid of bugs and crude edges, and identify some poor judgments we were making about how the interface should behave. We learned (once again) that four audio alternatives on an audio prompt were unacceptable, for example, "Press 1, listen again; 2, listen to another new message; 3, send a message; 4, hang up." In this example, we dropped the "4, hang up" alternative, but in other cases we had to reorganize the user interface.

These simulations were used to define the help messages. When participants were confused or stuck, they either volunteered what they wished they knew at that point (i.e., what information would help them), or we asked them. This methodology of asking people what they wish they knew when they are stuck has general value. The help messages were of two types: user requested and system initiated. If the user wanted help, he or she pressed the space (#) key, and a helpful message played out: "When in doubt, pound it out." At every point in the system, if the user paused beyond a specified time, we assumed that the user was having some difficulty. OMS automatically played a "time-out" message about what to do. The delay time for each of these messages was individually specified in the OMS interface description tables, and the initial values were determined in these simulations.

Unlike printed scenarios, live simulations or prototypes indicate how much a user must know to use the system. The designer cannot just slough off a usability problem with the platitude "We have a help system to take care of this." Simulations provide early tests as to whether the help system *actually* takes care of a problem—or introduces new problems since the user must now know how to use it also.

The need for consistent "escapes" was also identified in these simulations. "The escapes take subtly different forms," we wrote in our diary on February 5, "which the short printed scenarios that we have been using [described above] do not identify the need for." For example, in signing on (Figure 2), an Olympian might keypress a wrong (but valid) country code. This had jolting consequences. Instead of hearing "United States. Please keypress your last name," the user might hear "USSR" followed by a message in Russian that the American user did not understand. To address these problems, users pressed the backup (*) key, which "undid" their last action. Solutions like this require users to learn additional functions such as pressing the * key to backup. Designers often do not acknowledge the additional learning burdens that various help and error-correcting approaches place on learners.

Early Demonstrations. By February we were demonstrating audio versions of OMS to many people, with special emphasis on people from outside the United States who did not know much about computers. Here we received regular and strong emphasis to reduce function to the minimum. For example, in the family/friends interface it was recommended that we eliminate the ability for them to review a message before sending it. In the Olympian interface, it was recommended that we eliminate the ability to insert anywhere in a message. These functions, and others we dropped, were already smoothly accommodated into the user interface. The recommendations to drop them were made because additional function in a prompted interface comes at the price of additional prompts.

An Olympian on the Design Team. Throughout development we consulted with an ex-Olympian who competed for Ghana at the Olympics in Mexico City and Munich, and had participated in a large number of other international track meets. We received some insight into the content of messages Olympians might receive and send (e.g., he emphasized the joy of receiving "good luck" messages), how Olympians spent their time at previous Olympics (an important consideration in selecting kiosk sites, for example), and appropriate functions to provide. As it turned out, these periodic conversations tended to be brief, but very helpful. Perhaps that is the nature of "participative design."

Tours of Olympic Villages. The two Olympic villages were the large university campuses of USC and UCLA. Because of the many requests that vendors were making of the university people and the Los Angeles Olympic Organizing Committee (LAOOC), we were encouraged to stay away from these sites. We were told that we could look at maps or that the "appropriate people" could tell us what we needed to know. (But of course we would not know some of the important questions to ask until we saw the villages.) In the hectic world of system development, where key individuals have too much to do, this is, understandably, enough to discourage most system designers from going any further.

Walking around convinced us that it was not feasible to use classroom training for the Olympians. The campuses were large, and hilly (in the case of UCLA), and there was a chance the weather might be hot and smoggy. Living quarters were spread over many blocks, and large meeting rooms were unavailable.

Interviews with Olympians. We spoke with international competitors and officials from many countries, including some from previous Olympic games and likely candidates for the Los Angeles games. They encouraged us, telling us we were on the right track toward supplying a needed and very useful system for the Olympics. We learned how Olympians spend their time and how seriously the bus schedules to practice and competition can drive their personal schedules. We learned that more pointed systems questions could be asked of Olympians in later interviews—questions that we did not envision at first. For example, would Olympians want to send messages to Olympians in the other village? The answer to this question had serious systems implications.

Overseas Tests of the Family/Friends Interface. The audio prototype running in February gave us an opportunity to test, iterate, retest, etc., the family/friends interface from any telephone in the world (see Figure 1). There was no difficulty in getting participants with the right characteristics. We studied people in their homes and offices, at social gatherings, and at schools. They would read the "Family and Friends User Guide" (a pocket-sized card) and send a message to a specific person on the prototype system.

We also tested the family/friends interface from six South American countries to eliminate any unanticipated surprises due to overseas telephone systems. As so often happens, the emphasis in users' comments was on keeping things simple. This contributed to our eliminating several functions that were then running. Letting family or friends listen to a message they just recorded, or revise it before sending it if they had push-button telephones was dropped. We learned that the pamphlet should contain an example of how to use OMS, even though this would make it longer.

Hallway and Storefront Methodology. Visits to the villages startled us into the realization that we had to develop an Outreach Program—one that did not involve classroom training or was not people intensive in any other way. We decided that OMS kiosks would be the center of this program. They would give OMS a visual presence and provide a place for users to learn about OMS. They would reflect our philosophy of how people learn to do things. We believe people learn to do things not by reading about how to do them, but by observing and doing. Kiosks combined the traditional historic features of people gathering together around a bulletin board to get the news with today's electronic technology.

The kiosk (Figure 4) ultimately contained a CRT display of the names of Olympians having new messages, a push-button telephone, an instructional and entertaining videodisc of a mime demonstrating how to use OMS, and copies of the "Olympic Message System User Guide." The four-minute demonstration showed Olympians how to sign on, listen to a message, and send a message. They could see the demonstration and listen to it in their own language by pressing 1 of 12 buttons on the kiosk. (Readers can make a videotape copy of this demonstration by contacting us.) The user guides were also in 12 languages.

We started the design of these kiosks in early March 1984 following our return from visiting the villages. We put a 8-foot high, 45-inch diameter hollow cylinder in the front hallway of the Yorktown Research Lab in March. First, we simulated by pasting on this big cylinder CAD/CAM drawings of the two displays, the telephone, user guide holders, and instructional signs. Immediately we began to get comments and helpful suggestions from passersby, and their enthusiasm rubbed off on us. We were no longer just another five-person group. After a month we settled on the heights and locations of the displays, telephone, and user guide holders. This was done through several iterations of relocating the drawings and continuing to listen to comments. After these simulations we began the carpentry work of cutting holes in the prototype kiosk and making shelves. Hundreds of people viewed the (always interim) results and gave us useful comments. The wording of the English signs on how to use OMS and the translation into French were improved. Labeling the user guide holders in the wording and alphabet of the appropriate country, rather than in English, was suggested—for example, Deutsch rather than German. Aesthetic aspects were improved with suggestions, and a 72-inch crown was added to the kiosk for appearance sake (Figure 4).

We taped 12-foot wide, detailed CAD/CAM drawings of the kiosk on the wall in the hallway. These drawings elicited many comments also, often from craftsmen and construction workers. When it came time to manufacture kiosks, these plans and the working prototype were what the manufacturer followed.

By April we had a running OMS interfaced to the prototype kiosk. We put names of Yorktown people on the display in the hallway, together with a sign asking visitors whether their names were among those scrolling down the screen. People's comments helped determine the layout, scanning rate, and color of the display of Olympian names. The hallway kiosk provided an invitation for passersby to use OMS, which allowed us to receive even more comments. People volunteered to help us in other ways, for example, to do some initial language translation or spend the summer in Los Angeles.

Hallway methodology is an easy way to get participants for informal experiments. Besides being very useful, this methodology is exciting and personally rewarding. We wrote in our diary on April 4 that the "main feedback was how attractive, fun, useful" the kiosk and project are. "People really like looking for their names on the display. It gives our work an exposure and status out of proportion to only a five-person effort. It makes the project seem really important." Hallway methodology gives a project a visibility and existence that it would not otherwise have—especially in the early stages. This technique distinguishes a project from other projects. It accelerates beyond intuition the rate of progress. Other group members get a better feel for where their work fits in.

Yorktown Prototype Test. In preparation for a pre-Olympic field test, we conducted an intensive prototype test with about 100 participants. The use of prototypes can create several changes. Alavi [1], in analyzing 12 recent information systems projects that used prototyping, found that design managers felt it was harder to plan, control, and manage systems development when prototypes were involved, because they had to depart somewhat from fixed plans. At this point, OMS worked in four national languages. This test served mainly to debug the system and user interfaces. At the same time, it identified what we considered "trivialities," but in fact were not; for example, a system prompt we ultimately intended to change. It further identified what some of the help messages should be. It led to the tuning of the amount of time that should expire before a time-out message played out. It led us to create a "speech flow meter," whereby we could detect when a user, while recording a message, had stopped talking.

Win-a-Teddy-Bear Contests. We offered free coffee and doughnuts to anyone (of about 65 people) who would, for example, be the third person to change his or her password, or send the fifth new message to John Richards, or answer a message from Jim Schoonard. These tests were typically brief and done in a spirit of fun. We would quickly announce on everyone's time-sharing visual terminal who had won. If a person uncovered a particularly pernicious bug, we would find a suitable reward. Olympic pins were one of these.

Such tests heighten the general awareness of usability—for the developers, the participants (who in turn may become more conscious of usability in their own projects), and management. They can give management an early and quick view of likely customer reaction. They sure catch bugs.

Try-to-Destroy-it Tests. These tests were conducted periodically through the end of June, following a major modification to OMS. On one occasion we had 10 people in the same room, each using OMS from a different telephone while we watched the computer console. Just prior to the Olympics, we had computer science students from a local college dorm try to crash OMS in the evening. They needed no motivation beyond trying to bring down the system. In Los Angeles we once had 24 students call OMS at exactly the same moment, pressing keys in unison. All these tests contributed to the reliability of the system.

Having outsiders try to crash your system eliminates the unconscious tendencies of system designers to gently avoid the soft spots in their own systems while conducting such tests. Conducting these tests requires courage and humility because people will find problems with your system. But it is in this test arena that these battles should be fought—not subsequently on a customer's territory. As can be seen, our several types of iterative testing were informal, rather than controlled formal experiments [2, 5].

Pre-Olympic Field Test. By early April we had a user interface and the "Olympic Message System User Guide," which we felt were excellent, based on user tests using the methodologies already discussed. OMS was installed in Los Angeles to be used at a pre-Olympic event with competitors from 65 countries. We quickly learned, to our concern, frustration, and sadness, that our interface was not as good as we had thought. The problems were small, but cumulative. At the end of the five-day event, we had a list of 57 usability items that had to be addressed before the Olympics (see the sidebar, next page, for a discussion and solution to a few of these).

The study introduced us, firsthand, to an international reality that we intellectually knew existed, but had not personally experienced. We had already decided that OMS should work in more than the two official languages of the Olympics (English and French), but we now learned just how correct that judgment was. The delegations of competitors from Oman, Columbia, Pakistan, Japan, and Korea were unable to use our system, due to language barriers. Watching this helplessness and hopelessness had far greater impact than reading about it. It was embarrassing to us. Had we not been able to modify all usability aspects of OMS as a result of this international field study, we would not have been as successful at the Olympics.

Yorktown Final Prototype Test. In this test, we joined 2800 people to OMS. It was done primarily to check system reliability with a large number of users, with subsets assigned to different countries.

LAOOC Final Prototype Test. In parallel with the above test, we joined 1000 people to another prototype OMS running in Los Angeles. These people used OMS in their work for several weeks prior to the Olympics. This was particularly useful in learning how to interface OMS with the Los Angeles telephone network (which involved three telephone companies).

Hallway methodology gives a project a visibility and existence that it would not otherwise have—especially in the early stages. This technique distinguishes a project from other projects.

DISCUSSION OF THE PRINCIPLES

The Basic Principles Were Reinforced

The three basic principles worked. One must focus on users early to learn the type of system required. Empirical measurement and iteration are musts if a system is to be reliable, responsive, useful, learnable, usable, and desirable. Our experience is that anyone who has tried any of these principles believes in them afterwards. In contrast to the vast majority of systems on which designers have little to say about the methodologies used, designers actually boast about using this methodology. At the very minimum, by following these principles a reliable, responsive system can be achieved.

Sometimes we are asked when, in this method, iteration finally ends. One answer is that, with testable behavioral specifications, you know when you have reached your goal. We did not formally specify the criterion values for these specifications at the outset of OMS, which is not what we recommend to others. Another answer is that it never really ends—the test site just shifts. With many systems new releases are already being planned when the present one is announced.

Integrated Usability Design: A Fourth Principle

Based on our OMS experience, we raise "integrated usability design" to the level of a principle. There are two

A Field Test Identified Required Changes

In April 1984 we conducted a field study at a pre-Olympic event in California with international competitors from 65 countries. We discovered serious problems based on non-English-speaking users—problems we were aware could crop up, but whose seriousness we did not fully appreciate. The following example illustrates three points: (1) You cannot get it right the first few times, (2) different behavioral methodologies yield different types of information, and (3) integrative design is necessary to achieve good usability.

Prior to the field test, we felt we had a good Olympian sign-on sequence. It had gone through much behavioral analyses and several iterations of testing and redesign—all on English-speaking participants. The black type of Figure 5 shows the sign-on sequence as it was for this field test. Unfortunately, it did not work well. The parts of Figure 5 in color show changes that were made for the Olympics as a result of this study.

COUNTRY CODE

When a user called, OMS greeted and asked the user, first in English and then in French (the two official Olympic languages), to keypress his or her country code. Three problems came to light at this point. First, European users sometimes confused their Olympic country code with the concept of an international dialing code prefix. Inserting the phrase "three-letter Olympic" in the message, as shown in Figure 5, reduced this problem. This seemingly modest change had to be done in the interface, the help system, the reading materials, and the translations. Another help message was also added to reduce or eliminate this problem for the Olympics.

User: (Dial 8540.)
OMS: Olympic~~s~~ Message System.
Please keypress your three-letter Olympic country code.
~~Systeme de Messages Olympique:~~
~~Tapez le numer de votre pays, s'il vous plait.~~
User: U|S|A
OMS: United States. Les Etats-Unis.
Please keypress your last name.
User: ~~goul~~ G O U L
OMS: John Gould.
Please keypress your password.
User: 3|1|9
OMS: ~~Welcome to the Olympics Message System.~~
New Message from Stephen Boies.

FIGURE 5. Sign-On Sequence Used in the April Pre-Olympic Field Test, with Later Amendments in Color

Second, playing the first message in both English and French (see Figure 5) was too wordy. New users typically wait until an audio system becomes silent before beginning keypressing. They do not interrupt a message and start keypressing (as do experienced users). The switch in national languages was confusing. To remedy this for the Olympics, the first message played out in English only. Users could request the French version by pressing the asterisk (*) key. The user guide was suitably modified, and a mime demonstration and kiosk signs were developed to reflect this change. The solution to these two problems was possible because all aspects of usability were under one person's control and could be so integrated.

Third, the "time-out" help messages had been tuned for English. But, since the prompts were of different lengths in different languages (generally longer than in English), the tunings were good for English only. There was no way to make different tunings for different languages. So an algorithm was developed to start the time-out clock at the completion rather than the beginning of a time-out message, and this worked well for the Olympics. This was possible because the user interface, system functions, and system messages were each separate from each other.

NAME

There were five main problems when users were asked to keypress their last name. First, some Middle Eastern and Far Eastern users were not sure whether to keypress their first name or their last name. This was solved for the Olympics by emphasizing, with a picture in the user guide, in the time-out help messages, in the user-requested help messages, and in the mime's demonstration, that users should spell their names exactly as spelled on their badges. Again, this demanded an integration of four usability components—user

aspects to this principle: We believe that all usability factors must evolve together, and responsibility for all aspects of usability should be under one control. The sidebar provides an example of why integrated design is necessary.

Usability Factors Should Evolve Together. Usability is even broader than we had originally thought. Language translation of the user interface and reading materials was a much larger task than we envisioned (cf. [4] for a description of our behavioral approach to language translation). Another component has to do with user groups in addition to end users. In OMS there were several groups of operators, representatives, and runners who were responsible for day-to-day operations. Their jobs had to be defined and organized. For example, the system operator had to join 15,000 potential Olympians to OMS, which required recording their names in audio form. Assuming it took one minute to do this for each Olympian, this would require 30 person-days. Efficiency and correct pronunciation of names were critical. People fluent in all 12 languages needed to be selected and trained (see [4] for further discussion). Another usability component was the maintenance plan and people responsible for it, all of whom had to be recruited, trained, and supplied with reading materials. OMS had to be staffed by all these groups 24 hours a day.

Usually the major components of usability are developed sequentially, even though they jointly interact and affect each other. In OMS they were developed concurrently. We believe this was critical to our sucinterface, help systems, reading material, and training or outreach.

Second, in 5 of the 12 languages—Arabic, Chinese, Japanese, Korean, and Russian—user names had to be transliterated into the English alphabet. We observed in our April field test that some of these users were cavalier about how their names might be spelled on OMS. This problem was solved by referring them to the spelling on their badges, thus demanding an integration of the user interface, help systems, reading material, and demonstration.

Third, from the previous problem we anticipated that users might not know how to spell the name of an Olympian, especially from another country, from whom they had received a message and wanted to send a message in return (a very likely possibility we were told in interviews). We added a prompted "Reply" function (something we had previously dropped in the printed scenario stage) that eliminated the spelling requirement under these circumstances.

Fourth, users were required to press only enough characters in a last name to distinguish it from other Olympians from that same country. This was a feature that we had learned users wanted in ADS [8, 9]. In signing on, it caused serious problems in the field test. If a user pressed more than the minimum required characters in his or her last name, OMS considered them to be the first characters in his or her password (since the user's name had already been sufficiently specified from the system's point of view). Users became confused, particularly those who had no experience with "logging on" to a computer with the typical userid-password sequence. The problem was remedied for the Olympics by our creation of a "smart sponge," a part of the program that "soaked up" extra keypresses in a user's name. Once the user diverted from this string, OMS considered the characters to be the user's password. This algorithm worked better than earlier ones we developed based on interkeypress timings. The important metapoint here is that system resources must be used to solve usability problems.

Lastly, we had 4 national languages running for this field test. According to statistics from the Olympic Committee, about 80–90 percent of Olympians could be expected to speak 1 of these 4 languages. We decided that this was not enough, and went on to make OMS work in 12 national languages. This had implications for several aspects of usability—user interface, reading materials, and mime demonstration.

PASSWORD

The key problem here was that users did not always know their password. We solved this for the Olympics by initially assigning as a password the last three digits of each user's badge number. The mime demonstration and the user guide emphasized this point. Users could change their passwords, as shown in the user guide.

Some non-English-speaking users told us that it was confusing to answer some prompts by entering characters (i.e., country code and name) and all others by entering digits. "You have to read the telephone keys differently," one explained. Although we kept this approach, we did change the typography of the user guide to help reduce this confusion. For example, as shown in Figure 5, we switched to uppercase characters (because the telephone keys have these) to show that the user was keypressing characters, and we put spaces between the characters, as several users suggested.

ADDITIONAL USABILITY CONSIDERATIONS

This field study convinced us that we needed an enhanced Outreach Program and an even more streamlined user interface than we had already developed. We settled on the notion of using a mime demonstration for instruction. The contents of the demonstration particularly addressed the sign-on sequence. We added the concept of a Help Line, with the ability to add on OMS to a conversation between a user who needed help and a helper. We emphasized this in the signs on the kiosk and in the user guide. This Help Line was used frequently in the Olympics, and it guaranteed that, no matter where an Olympian was, he or she could get help in sending or receiving messages. In streamlining the user interface further, we dropped the ability for a user to send the same message to more than one person. Only through a design philosophy where all aspects of usability are integrated could we have addressed the problems illustrated in this example so fast and so successfully. Simply modifying a couple of messages in the already existing user interface or a couple of lines in the already written user guide would not have done it.

cess. In this parallel evolution, all components were refined based on the same methodologies of testing and redoing.

Within each main component, usability is made up of many details. A few examples in OMS included the decision to use original or established music in the mime's demonstration; the balance between entertainment value and instructional value in the mime's demonstration; the type of paper to be used in the "Olympic Message System User Guide" (glossy paper might be subject to glare or might be hard to write on, whereas a light bond might disintegrate in sweaty back pockets); and the placement of page breaks in the example-of-use scenarios in the user guides. Once all of this was decided for English, the other 11 languages had to be considered. For example, which way should the Arabic user guide fold? Also, other features of the kiosks such as air conditioning, rain proofing, and the type of phone cord to use all had to be decided.

Responsibility for Usability under One Focus. It is impossible for all aspects of usability to develop in an integrated way when responsibility for it is spread over several groups, some of whom do not begin their work until others have nearly completed theirs. There is too much to remember, too much to negotiate, and too much to do. If responsibility for usability is fractionated, even the simplest changes are routinely difficult and require negotiation. It is vastly simpler to make changes yourself than to request others to make them, and live with the uncertainty of which ones will be made. Iterative design means lots of drudging work. No secretary would willingly retype the reading material as often as we did. How could a designer explain to anyone that he or she made so many mistakes that a user guide had to be modified 200 times?

Why Could We Follow the Principles?

Why were we able to follow these principles when few other designers do and when management often says it is impossible? First, we had good tools. The layered ADS architecture on which OMS was built was designed for iterative design (see [12]). There was a separation of user interface and function. The table-driven ADS user interface was easy to change. The VM "simulator," used two years earlier, was relatively easy to update.

Second, we committed ourselves to follow these principles, using OMS as a test case. We deeply believed that usability should drive system design. Arguments about how to proceed were settled by an appeal to the principles (e.g., see [4]).

Third, we were a small communicative group, and this reduced the need to formalize and freeze important usability characteristics very early. It made living with change possible and relatively easy.

We can exclude some reasons. The complexity of the system and probably the knowledge of the application do not seem to determine whether one can follow these principles. It is not sufficient to have human-factors people on the project. We can observe in our own backyard projects that do not follow these principles, even though they have human-factors people in key roles.

The data of this study are correlational—one system was developed with one design method. In this sense they are similar to other valuable case histories discussing systems development methodologies; for example, IBM's ADS [8], Tektronix's Graphic Input Workstation [16], Boeing's banking terminal [6], Digital Equipment Corporation's VAX Text Processing Utility [7], Xerox's Star system [14], Apple's Lisa system [17], and Swezey and Davis's [15] report on trying to apply existing human-factors guidelines in developing a graphics system.

Our study does not prove that the design principles used here are better than others. To do this would require a comparative study of several design methodologies (as the independent variable) with all other variables held constant. This scientific approach is possible only with problems of a much smaller scale (see [3] for comparison of two design approaches, prototyping versus specifying).

We have tried to assess honestly the value of the design principles used here. In our opinion the principles were necessary, but not sufficient, for the success of OMS. There were three other general factors that also contributed critically: first, the people themselves. We had a powerful, dominant leader, Stephen J. Boies, who understood most everything and drove everything. We were the ones who had done much of the design and programming of the prototypes (see [9]) that became IBM's ADS product—the system OMS was built upon. We had the right skill mix—of telephony, systems, application, and human-factors expertise. We had tremendous self-imposed pressure for success. We had no intention of failing—but we feared it (see [4] for a description of some of our fears and conflicts, and what it was like to work at the Olympics).

Second, we had outstanding support—both people and facilities. Without the variety of skills and facilities at Yorktown, OMS would not have been as good. When we needed somebody, we got the right person—and right away.

Third, the Olympics itself contributed. It greased the wheels in getting us what we wanted. The absolute certainty of the begin and end dates, and the certainty that we would be completely finished by the end of the summer, allowed us to work at the pace we did. In retrospect, we have not been able to identify another event—personal, national, or international—that generates comparable enthusiasm and inspiration.

What We Would Do Differently. In retrospect, there is no major aspect of OMS that we would change. Elsewhere, we have mentioned five minor things we would do differently [4].

What Is Exportable?

We learned four general points that can be applied to other systems. First, the principles of design are exportable. They are needed so you know what you are doing. In telling developers about the success of following these principles in developing OMS, we have occasionally noticed an attitude to trivialize OMS. It is worthwhile to remember that, although OMS was built by a few people in a short time, it was a large system (network of over 35 computers), contained a significant new function, and worked reliably and successfully. Furthermore, it was a high-risk system in the sense that it was very visible, handled sensitive information (personal communications), was potentially subject to abuse and sabotage, and could have failed in many different and public ways.

We committed ourselves to follow these principles, using OMS as a test case. We deeply believed that usability should drive system design.

Second, the notion of layered system design is exportable. This is needed so you can make changes.

Third, the concept of tools for interface designers, as implemented in the Voice Toolkit or User Interface Management System used here [12], is exportable. They allow implementers to code and debug much more rapidly. They allow human-factors people to compose, test, modify, refine, and control the user

We had the right skill mix—of telephony, systems, application, and human-factors expertise. We had tremendous self-imposed pressure for success. We had no intention of failing—but we feared it.

interface—without having to be systems programmers. The mind-set remains in the right perspective. Tools enhance individual productivity and, more importantly, allow you to have smaller groups. Separating the organization and details of the user interface from the functions it calls provides a useful division of labor and adds greatly to the ultimate system goodness.

Fourth, the commitment to living in a sea of changes—and making them—is exportable. You need this to make a good system happen.

CONCLUSIONS

There were remarkable aspects to OMS: Planning and development were done over a short time by a small number of people, and second, it worked well and was a success. OMS was used more than once a minute, 24 hours a day during the Olympics. The project demonstrated that behavioral principles of design could be, and were, followed.

Rather than impede the development process, as is sometimes suggested, following these principles speeds up the development process by identifying right and wrong directions early, and by making change easy. Extra effort in the early stages, which these principles seem to require, leads to much less effort later on and a good system at the end. The project demonstrated that following these principles can be done, does not take too long, and does not cost too much. The principles made possible an integration of all aspects of usability. They led to a reliable, responsive, easy-to-learn system containing the right functions.

Acknowledgments. We received help from many people at IBM's Thomas J. Watson Research Center in Yorktown Heights and at IBM Corporate Headquarters. We especially thank Bill Bennett, Nils Bruun, Rich Finn, Walt Gray, Marilyn Hoppe, Ann Hubby, Linda Klapp, Bill Lewis, Clayton Lewis, Peggy Lohr, Jack McMahon, Angela Minuto, Mike Odierna, Michael O'Sullivan, Michael Starks, and Linda Tetzlaff. The IBM Los Angeles Olympic Project Office was managed by Brian King. Sue Blair had responsibility for OMS. We thank them, and their directors Paul Barton and Gene Fairfield, who provided so much help and expertise when it was needed. Jack Kesselman and Cliff Lau very effectively managed OMS installation and operations in the villages.

REFERENCES

1. Alavi, M. An assessment of the prototyping approach to information systems development. *Commun. ACM 27*, 6 (June 1984), 556–563.
2. Bennett, J. Managing to meet usability requirements: Establishing and meeting software development goals. In *Visual Display Terminals: Usability Issues and Health Concerns*, J. Bennett, D. Case, J. Sandelin, and M. Smith, Eds. Prentice-Hall, Englewood Cliffs, N.J., 1984, pp. 161–184.
3. Boehm, B.W., Gray, T.E., and Seewaldt, T. Prototyping versus specifying: A multiproject experiment. *IEEE Trans. Softw. Eng. SE-10*, 3 (1984), 290–302.
4. Boies, S.J., Gould, J.D., Levy, S.E., Richards, J.T., and Schoonard, J.W. The 1984 Olympic Message System—A case study in system design. Res. Rep. RC-11138, IBM, Yorktown Heights, N.Y., 1985.
5. Bury, K.F. The iterative development of usable computer interfaces. In *Proceedings of INTERACT 84—First IFIP Conference on Human-Computer Interaction* (London, Sept. 4–7). Elsevier Science Publishers, Amsterdam, 1984, pp. 343–348.
6. Butler, K.A. Connecting theory and practice: A case study of achieving usability goals. In *Proceedings of CHI 85 Human Factors in Computing Systems* (San Francisco, Calif., Apr. 14–18). ACM, New York, 1985, pp. 85–88.
7. Good, M. The use of logging data in the design of a new text editor. In *Proceedings of CHI 85 Human Factors in Computing Systems* (San Francisco, Calif., Apr. 14–18). ACM, New York, 1985, pp. 93–98.
8. Gould, J.D., and Boies, S.J. Human factors challenges in creating a principal support office system—The speech filing system approach. *ACM Trans. Off. Inf. Syst. 1*, 4 (Oct. 1983), 273–298.
9. Gould, J.D., and Boies, S.J. Speech filing—An office system for principles. *IBM Syst. J.* 23, 1 (1984), 65–81.
10. Gould, J.D., and Lewis, C.H. Designing for usability—Key principles and what designers think. In *Proceedings of CHI 83 Human Factors in Computing Systems* (Boston, Mass., Dec. 12–15). ACM, New York, 1983, pp. 50–53.
11. Gould, J.D., and Lewis, C. Designing for usability: Key principles and what designers think. *Commun. ACM 28*, 3 (Mar. 1985), 300–311.
12. Richards, J.T., Boies, S.J., and Gould, J.D. Rapid prototyping and system development: Examination of an interface toolkit for voice and telephony applications. Res. Rep. RC-11433, IBM, Yorktown Heights, N.Y., 1985.
13. Rubenstein, R., and Hersh, H. *The Human Factor—Designing Computer Systems for People*. Digital Equipment Press, Burlington, Mass., 1984.
14. Smith, D.C., Irby, C., Kimball, R., and Verplank, B. Designing the Star user interface. *Byte 7*, 4 (Apr. 1982), 242–282.
15. Swezey, R.W., and Davis, E.G. A case study of human factors guidelines in computer graphics. *IEEE Comput. Graph. Appl.* 3, 8 (Nov. 1983), 21–30.
16. Weiner, H. Human factors lessons from the design of a real product. TEK Tech. Rep., Tektronix, Wilsonville, 1984. (Can be obtained from the author at Tektronix, Box 1000, Wilsonville, Oreg., 97070.)
17. Williams, G. The Lisa computer system. *Byte 8*, 2 (Feb. 1983), 33–50.

CR Categories and Subject Descriptors: D.2.2 [**Software Engineering**]: Tools and Techniques—*user interfaces*; D.2.m [**Software Engineering**]: Miscellaneous—*rapid prototyping*; H.1.2 [**Models and Principles**]: User/Machine Systems; H.4 [**Information Systems Applications**]; H.4.3 [**Information Systems Applications**]: Communications Applications—*electronic mail*

General Terms: Human Factors

Additional Key Words and Phrases: Human-computer interaction, Olympics, voice mail

Authors' Present Address: John D. Gould, Stephen J. Boies, Stephen Levy, John T. Richards, and Jim Schoonard, IBM Research Center, Box 704, Yorktown Heights, NY 10598.