INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION, *14*(3&4), 447–462

Extending the Usability of Heuristics for Design and Evaluation: Lead, Follow, and Get Out of the Way

Robert J. Kamper
IBM

User-centered design practitioners have often relied on discount usability engineering methods using heuristics. Top 10 lists of design and evaluation heuristics have proliferated during the 1990s, leading to a plethora of heterogeneous heuristic guidelines for a multitude of user interfaces. A simple, unified set of heuristics that might be applicable across different technologies, understood between disciplines, and equivalent to metrics used to measure ease of use can be beneficial.

Lead, follow, and get out of the way (LF&G) theorizes that the optimal human–computer interaction (HCI) experience is analogous to a facilitative learning relation—Like a good teacher, mentor, or coach, the usable user interface leads the user to successful completion of tasks and goals; follows the user's progress and provides appropriate feedback and information when needed; and gets out of the way of the user to allow efficient and effective completion of tasks as the user attains mastery of the system, its concepts, and operations. A set of 18 heuristics grouped under the 3 general principles of the title are provided for use across the tasks of design guidance, development refinement, and end-user evaluation of computer systems.

This article provides background into the development of LF&G, case studies on its use in real-world product design and development, and directions for further research needed to develop this approach into an empirically based method for defining, describing, designing, and predicting the ease of use of interactive human–computer systems.

The author acknowledges the cooperation of many IBM colleagues during the development and implementation of the LF&G principles and practices to this date. Among those whose use of the forms and feedback related to practical matters has assisted in developing the forms and demonstrating the usefulness of the approach are Doug Bloch, John Hilburn, and Theresa Deron. Thanks are also extended to Alex Gamez, Jeff Kline, Ed Kunzinger, and Karel Vredenburg for their encouragement and support of the development and dissemination of LF&G theory and practice within the framework of IBM's Integrated Product Development and User-Centered Design processes. Further thanks are due to the anonymous reviewers of the original version of this material, whose frank feedback and honest criticism have contributed to the quality and clarity of this article. Any remaining confusion, lack of coherence, or inclusion of colorful metaphors are solely the responsibility of the author.

Requests for reprints should be sent to Robert J. Kamper, 11400 Burnet Road, Austin, TX 78758. E-mail: rkamper@us.ibm.com

1. INTRODUCTION

Usability is now recognized as an important factor in the success of interactive computer systems (Butler, 1996). To ensure product usability, user-centered design (UCD) has become an accepted method in the hardware and software design and development process. Within UCD, heuristic evaluation has become a widely used tool. The popularity of heuristic evaluation as a discount usability engineering evaluation method can be attributed to several factors:

- It can be conducted relatively cheaply using only experts and lists of heuristics, in contrast to usability laboratory evaluations, which typically require specialized facilities and skills resources in addition to the software or hardware being evaluated.
- It can be conducted relatively quickly in contrast to formal usability laboratory evaluations, which require scheduling of facilities and participants. Also, it can be repeated quickly during the development process to obtain multiple evaluations and to assist in prioritizing development fixes and schedules.
- It is an easy method to use in comparison to formal usability laboratory testing, which usually assumes the rigorous practices and methodology of experimental psychology and quantitative evaluation.
- When combined with a first task-oriented pass, followed by a review involving a list of heuristics in a heuristic walkthrough, serious usability problems are uncovered and the technique is valid, reliable, and thorough (Sears, 1997).

2. BACKGROUND

Research into the comparative effectiveness and efficiency of usability engineering methods has resulted in a general conclusion that both heuristic evaluations and user testing have different strengths and provide qualitatively different results that contribute to improving usability. Typical findings recommend using expert heuristic evaluations early in the design process, followed by user testing later in the process, when working prototypes are available (Doubleday, Ryan, Springett, & Sutcliffe, 1997; Jeffries, Miller, Wharton & Uyeda, 1991; Kantner & Rosenbaum, 1997; Nielsen & Phillips,1993; Savage,1996).

One line of research has been the exploration of optimal methods for the implementation of heuristic evaluations. An evaluator effect has been found due to evaluators' length of usability discipline experience as well as the level of domain-specific knowledge. For this reason, heuristic evaluation by a single evaluator is not reliable. At least three to five expert evaluators per evaluation are recommended to achieve optimal effectiveness and efficiency, as well as to defend against threats to the validity of the findings (Jacobsen, Hertzum, & John, 1998; Nielsen, 1992, 1993; Nielsen & Molich, 1990).

Another trend has been the proliferation of domain-specific heuristics. Some Web design guidelines and heuristics (e.g., IBM Ease of Use, n.d.) may or may not be related to the well-known set of nine heuristics defined by Nielsen and Molich (1990). Others begin with the Nielsen and Molich list and add additional heuristics based on the specific domain involved (Borges, Morales, & Rodriguez, 1996; Köykkä et al., 1999). This has lead to a diversity of guidelines for a multitude of contexts.

Other investigations have sought to improve on the comprehensiveness of the heuristics used and the effectiveness of the techniques employed in expert evaluations (Muller, Matheson, Page, & Gallup, 1998; Nielsen, 1994; Sears, 1997).

3. LIMITATIONS OF CURRENT PRACTICES

Like many solutions, the widespread use of heuristic evaluations as a cost-effective usability engineering practice has created new problems. The proliferation of heuristics and guidelines, each set customized for a specific context, in itself presents a usability problem for usability practitioners and their customers. For usability practitioners, the task of becoming proficient in a wide range of contexts requires the learning of disparate guidelines—Heuristics and guidelines for one context might not be applicable to a different context. Indeed, a guideline for one context might contradict a guideline for another context. Customers in one knowledge domain may not consider guidelines developed in a different domain applicable to the current context.

Limitations with current ease of use, UCD, and usability engineering practices include

- Usability evaluation is oriented toward fault finding, instead of goal achievement.
- Usability guidelines for software, hardware, documentation, and the Web are disjointed.
- A plethora of guidelines exist with no comprehensive underlying basis.
- Design and evaluation heuristics are incongruent.

One limitation with the current use of heuristics in usability engineering is that heuristics are intended primarily to identify problems in design instead of measuring achievement of design goals. An approach that leads to success with a minimum of problems, predicts which problems are likely to occur, and helps to provide solutions to those problems so that users are able to successfully achieve goals during initial and subsequent product encounters is needed.

4. NEED FOR NEW DIRECTIONS

So far, the primary thrust for improving heuristics has been to enhance the ability to explain usability problems and to find new problems (Nielsen, 1994). A comprehensive set of heuristics that can be used for both designing usability into a system and for measuring the effectiveness of the design has not previously been proposed. Such a set of heuristics might be useful for setting usability and ease of use goals, might provide design guidance, and can measure ease of use at the end of the design and development cycle. A simple, unified set of heuristics might be beneficial to practitioners. To be adopted, it has to be

- Applicable across different technologies.
- Applicable across different contexts and domains of knowledge.
- Understood between multiple disciplines.
- Equivalent to the metrics used to measure ease of use.

The model presented here is an effort to meet that need in a clear, easily learned, easily practiced, and effective manner. Building on the foundation established by Nielsen (1994), the sets of design guidelines and usability heuristics embodied by this method ought to be familiar to practitioners in the field. When used with both a task-based and a cognitive walkthrough (Sears, 1997), the model and method are hoped to prove valid for use, reliable over time, evaluators, and contexts, and comprehensive in scope.

This article presents three design principles that are comprehensive in scope and within which all design and evaluation heuristics can be categorized. These three principles provide the base that hopefully will enable not just the explanation of usability problems but also the prediction and prevention of problems during the design phase of development.

5. LEAD, FOLLOW, AND GET OUT OF THE WAY (LF&G)

Three design principles provide a comprehensive structure within which iterative design and evaluation activities are conducted to ensure ease of use in the final product. These principles are

Principle 1: Lead the user to successful achievement of tasks and goals.
Principle 2: Follow the user's progress and provide information and support as needed.
Principle 3: Get out of the way to allow users to perform tasks efficiently and effectively.

Simplified into a mnemonic, LF&G theorizes that the optimal HCI experience is analogous to a facilitative learning relation—Like a good teacher, mentor, or coach, the usable human–computer interface leads the user to successful completion of tasks and goals; follows the user's progress and provides appropriate feedback and information when needed; and gets out of the way of the user to allow efficient and effective completion of tasks as the user attains mastery of the system, its concepts, and operations.

The order of the principles follows the assumed order of priority—It is more important to successfully lead the user to success in the first place (without going off task) than it is to provide additional information and support during the process of completing a task. Likewise, it is assumed that it is more important to provide additional information, guidance, or support in the middle of a task than it is to get out of the way of the user to allow the user to complete the task more quickly (e.g., getting in the user's way and forcing a conscious response to a confirmation of a command to delete all files instead of simply deleting all files immediately). Therefore, the order of the mnemonic follows the order of importance within the model, and can help in communicating priorities of findings and managing progress toward completion of a project.

The LF&G approach includes a set of 18 heuristics grouped under the three general principles of the LF&G title, provided for use across the tasks of design guidance, development refinement, and end-user evaluation of computer systems (see Table 1). It includes the development of heuristics checklists and evaluation forms for use as aids and tools during requirements definition, design, development, and evaluation phases of computer system development (Kamper, 2001). Although still in the process of being refined and developed, the LF&G materials and approach are offered as a theoretical and practical base with which to extend the usability of heuristics for design and evaluation in the development of human–computer systems, subject to validation and refinement by traditional research practices.

6. UNDERSTANDING THE LF&G PRINCIPLES

6.1. Principle 1: Lead the User to Successful Achievement of Goals

The interface must be obvious enough that the typical user can tell what tasks and goals can be accomplished through the system, simple enough that the first-time user can be successful achieving goals, and robust enough that the experienced user can perform tasks with a minimum of errors and a maximum of speed and economy of effort. When the user interface violates this principle, no amount of help and documentation inserted into the interface in an attempt to provide corrective (follow the user) assistance can overcome the misdirection.

A well-known example of a violation of this principle occurred in the November 2000 U.S. presidential election ballot in Palm Beach County, Florida. The now-infamous butterfly ballot presented the candidates in a list that spanned two pages, with a single column of punch holes, now commonly referred to as *chads,* in the center between the two pages. The incongruence of the choice list and the corresponding chads, as well as the incongruence of the order of reading contributed to

Table 1: The Lead, Follow, and Get Out of the Way Principles and Supporting Heuristics

1. Lead the user to successful achievement of goals.
   - 1.1. Make interface functions obvious and accessible to the user.
   - 1.2. Prevent the possibility of errors on the part of the user—Hide, disable, or confirm inactive or potentially destructive actions.
   - 1.3. Make labels and names distinct from one another—Avoid ambiguity and confusion.
   - 1.4. Provide clear, concise prompts in users' own terminology and language.
   - 1.5. Provide safe defaults for inputs—recognition, not recall, of information.
   - 1.6. Support the natural workflow or taskflow of the user.
2. Follow the user's progress and provide support as needed.
   - 2.1. Provide feedback on all actions.
   - 2.2. Provide progress indicators when appropriate due to length of time elapsed during action.
   - 2.3. Provide error messages that offer solutions to problems.
   - 2.4. Provide feedback on successful completion of a task.
   - 2.5. Provide ability to save input as template in future, record macros, customize preferences, and so forth.
   - 2.6. Provide goal- and task-oriented online help and documentation.
3. Get out of the way to allow the user to perform tasks efficiently and effectively.
   - 3.1. Minimize the number of individual actions needed to perform a task.
   - 3.2. Maintain consistency and adhere to platform conventions and user interface standards.
   - 3.3. Allow the user to maintain control—Provide undo, redo, and user exits.
   - 3.4. Provide an aesthetic and minimalist design—Shield user from minutiae unless desired by user.
   - 3.5. Provide for multiple skill and task levels.
   - 3.6. Provide shortcuts.

the confusion of some voters. The sequence of choices was A, B, C, D, E, F, G, H; the sequence of punch holes was A, E, B, F, C, G, D, H. Instead of leading voters to a successful completion of tasks and goals, it presented a usability challenge to all voters except those whose choice was option A on the list of candidates.

6.2. Principle 2: Follow the User's Progress and Provide Support as Needed

The second principle, for optimum benefit, requires a hierarchical task analysis, preferably occurring early in the design process. This analysis must identify the information and support needed by users at various points in the completion of tasks and goals. This can be used to present the right information to the user at the right time to assist the user in making the right choice, resulting in achieving the desired goal without error. Lacking this analysis, the LF&G heuristics may be used to identify those points in the interface that are lacking in sufficient information or affordances to enable the user to find the paths through the available tasks.

Design guidelines and usability heuristics typically address this principle vaguely or indirectly, with generalities such as, "provide feedback," "good error messages," or "help and documentation," relying on the knowledge and experience of the expert evaluator to know what to provide and how to provide it. LF&G expands on these topics to provide more specific and explicit guidance at the heuristic level for less experienced evaluators.

In the butterfly ballot example already noted, two different attempts to follow the user and provide support can be identified. First, the ballot itself included arrows pointing toward the chads associated with the choices in an attempt to compensate for the violation of the lead principle. Second, during the election, the county supervisor of elections sent out a memo (help and documentation) directing poll workers to remind voters to punch out the chad next to the arrow for the intended candidate.

This illustrates why error messages and help and documentation are often insufficient to overcome a failure to lead users to successful completion of tasks and goals due to an initial bad design. By focusing on errors and not on goals, the intended help often fails to provide the support needed and cannot overcome the initial misdirection.

6.3. Principle 3: Get Out of the Way to Allow Users to Perform Tasks Efficiently and Effectively

The third principle is to simplify the interaction so that users can perform tasks with the fewest steps and in as little time possible, even if it requires the creation of separate interfaces for different users and contexts. In practice, this is often implemented in the form of shortcuts for experienced and knowledgeable users, but it must also apply to the novice user and the occasional user as well.

Let us return to another voting example. A ballot used in Williamson County, Texas, during the November 2000 election did not receive the notoriety of the Palm Beach County ballot. One reason for the lack of attention given to this Texas ballot was the use of a better technology that resulted in fewer errors—an instance of getting out of the way of the user by reducing the effort and skill needed to create a scannable vote. The other reason for inclusion here is that the Williamson County ballot began with an option to vote a straight party ticket—the voting equivalent of one-click purchasing on the internet. Instead of forcing the voter to spend time and effort to identify and vote for each individual office on the ballot, the voter was able to fill in a single circle on the ballot and effectively cast votes in presidential, congressional, state, and local elections by party affiliation. The voting system effectively got out of the way of the voter whose intention was to vote by party lines, regardless of the party. In this instance, the advantage of efficiency and effectiveness was provided to both the expert voter with extensive knowledge of the issues as well as the voter whose knowledge of the issues was limited.

As this example shows, use of the LF&G approach might have easily identified the lead violation during design of the butterfly ballot. Because the intention of the butterfly design was to make it easier for elderly voters to read the ballot, it is clear that at least informal usability testing and customer input had already taken place, and the design was an attempt to correct the identified usability problem (a lack of readability due to small lettering). A more appropriate follow design point might have been to make sure voters were advised to punch the chad out of the ballot sheet completely, and to provide visuals of adequate and inadequate punches so the voters could compare their ballots against a model prior to casting the ballot into the ballot box. Finally, adoption of a single punch for a party line vote would have improved usability for those voters choosing to vote in that manner.

The now infamous butterfly ballot was a design change in reaction to user testing of previous ballots, an attempt to improve usability based on user feedback and complaints. Additional usability testing might have resulted in improvement, or might have resulted in the creation of new usability problems. It is believed that the LF&G approach might have provided better design recommendations, identified possible problems, and provided an advantage that current usability heuristics and evaluation practices lack.

7. BENEFITS OF LF&G

Some of the specific benefits that the LF&G approach has provided in practice thus far, while still in development include

- LF&G heuristics provide design direction. A butterfly ballot would not have been an LF&G design solution to voting usability problems. In addition to finding problems, the organizing principles of LF&G help to focus on the design solutions and facilitate communication with developers. This can be useful even when an existing design is being evaluated instead of being developed from the ground up.
- LF&G provides an embedded structure and priority for reporting and addressing problems. Under LF&G, it is more important to lead all users to successful completion of goals than it is to provide additional support along the way. In general, it is more important to provide contextual support along the way than it is to get out of the users' way. Again, this helps to facilitate communication with developers.
- The LF&G mnemonic is easy to remember and is easy to explain and communicate to non-usability practitioners, such as developers, programmers, managers and customers.
- The LF&G approach is easy for experienced usability, human factors, and HCI design staff to pick up and use. The 18 individual heuristics categorized in groups

of six under the three LF&G principles are gathered from the extant literature and include all 10 of Nielsen's (1994) refined set. The supporting heuristics are both validated within the literature and organized in a manner that helps to guide designers in creating usable designs.

- These principles and heuristics can be used for organizing and structuring rating evaluations and reports. This makes reporting easy to produce by evaluators, easy to understand by developers, and easy to use in support of project management concerns.

Although the necessary research has yet to be done, the following possible benefits seem worthy of investigation:

- These principles and heuristics can be used for the instruction of, and learning by, student usability evaluators. Whether the LF&G organizing principles provide any measurable advantages over current approaches in quickness or quality of learning remains to be determined.
- The use of six supporting heuristics within each of the three principles provides a structure for an evaluation instrument that should be reasonably robust. (See Lewis's [this issue] discussion of a similarly structured questionnaire that has been used for usability evaluations).
- If ratings obtained through LF&G evaluations can be correlated with usability satisfaction questionnaire results, it would enable practitioners to estimate, if not predict, customer satisfaction with general release products earlier in the development cycle.
- The LF&G method provides a ready-made division into three areas. Typically, heuristic evaluations are conducted by three or more trained evaluators. Whether the efficiency and effectiveness of heuristic evaluations can be enhanced or maximized by assigning each reviewer to a specific area, instead of consolidating several overall evaluations, is a research question worth investigating.

8. LF&G AND NIELSEN'S TOP TEN LISTS

Table 2 presents Nielsen's (1994) sets of heuristics and design guidelines in the context of the LF&G heuristics, along with the other lists of heuristics cited in the same article. In Column 1 are the LF&G principles and numbers representing the associated heuristics listed in Table 1. Columns 2 and 3 represent the two top 10 lists from Nielsen's (1994) article with the letters and numbers of the heuristics as listed under the columns to the right. For instance, the lead heuristic 1.1 is "Make interface functions obvious and accessible to the user." Both of the first two Nielsen (1994) top 10 lists refer to Item 1 in Polson & Lewis' 1990 list of heuristics, represented as *Po1.* The Po1 notation refers to Polson's first heuristic—"Make the repertoire of available actions salient." The next six columns refer to the lists of heuristics cited in Nielsen's (1994) article, including the Nielsen 1990 list. The numbers under each column refer to the items within that specific list and are listed in the rows corresponding to the LF&G heuristics itemized in the first column. This table shows the frequency in which the LF&G principles and heuristics and related concepts appear in earlier sets of guidelines. One conclusion that can be drawn from this table is that the LF&G heuristics have strong face validity. Another is that the LF&G heuristics encompass both those heuristics that explain a majority of usability problems

Table 2: LF&G Heuristics in Relation to Other Sets of Design and Evaluation Heuristics

1 *Lead, Follow, and Get Out of the Way*	*2* *Top Ten Heuristics to Explain All Usability Problems*	*3* *Top Ten Heuristics to Explain Serious Usability Problems*	*4* *Nielsen Top Ten*	*5* *Star User Interface*	*6* *Polson*	*7* *Carroll*	*8* *Macintosh Human Interface Guidelines*	*9* *Sunsoft*
1. Lead	—	—	—	3	2	14	—	—
1.1	Po1	Po1	—	—	1	1, 11	5	1
1.2	Mac8	Mac8, Ni9	9	—	6	4, 12	8	22
1.3	—	Po5	1	4	5	3	9	10
1.4	Ni2, Sun18	—	2	—	—	7	1	2, 3, 18
1.5	St2	St2	3	2	—	16	3	13
1.6	—	St1	—	1	—	2, 19, 5	—	8, 9, 11
2. Follow	—	—	—	—	—	—	—	—
2.1	Mac7	Mac7, Sun5	5	—	3	10, 13, 15	2, 7	4, 5
2.2	—	—	—	—	—	—	—	—
2.3	Ca18	—	8	—	—	6, 9, 18	—	—
2.4	—	—	—	—	—	17	—	—
2.5	—	—	—	8	—	—	—	—
2.6	—	—	10	—	—	16, 17	—	—
3. Get Out of the Way	—	—	—	—	8	—	11	16
3.1	—	—	—	6	—	8	—	12, 15
3.2	Ni4	Mac4, St7	4	5, 7	—	—	4	19
3.3	—	—	6	8	4	—	6	7, 23
3.4	Mac10	—	—	—	—	—	10	6, 20, 24
3.5	—	—	—	—	7	—	12	17
3.6	Ni7	—	7	—	—	—	—	14, 21

Note. This table provides the LFG heuristics in the first column and corresponding heuristics in the columns to the right, using the sets of heuristics cited in Nielsen (1994). Numbers that follow the abbreviations for names indicate the number of the heuristic developed by that group or individual. L&FG = Lead, follow, and get out of the way; Po = Polson; Mac = Mac HI; Ni = Nielson; St = Star UI; SUN = Sunsoft; Ca = Carroll.

as well as those that explain the most serious usability problems. We think that LF&G extends these two top 10 lists to enable us to explain all usability problems.

It should be noted that the grouping of the Nielsen (1994) lists and the other sets of design and evaluation heuristics in correspondence with the LF&G set is solely my responsibility, and reasonable persons might discuss and debate the categorizations as well as the level of granularity used. For instance, LF&G 2.2 and LF&G 2.4 provide separate heuristics, which those expert in the art of usability evaluation might identify as subsumed under the *Ni5* heuristic "Feedback: The system should always keep users informed about what is going on, through appropriate feedback within reasonable time." I simply judged that providing feedback on successful completion of a (user) task is a qualitatively different guideline than providing progress and busy indicators during (computer) processing times, and therefore, separate heuristics are provided in the LF&G set. Likewise, providing feedback on all actions is deemed a separate concept in the LF&G set. Each of the source heuristics was used only once, corresponding to the LF&G heuristic that was most equivalent to the source heuristic's main concept. Where LF&G heuristics represent concepts that might be secondary within the source concept, the source concept is not listed a second time. This was done more for the sake of simplicity than to claim that the source heuristic did not address the concept.

It should also be noted that the listing of individual heuristics under LF&G is a synthesis of prior work, and that the number of heuristics grouped under the three principles is an arbitrary grouping by one person. Further validation of my choices and refinement of the appropriate heuristics could be an interesting piece of work. The determination of six heuristics under each principle was also arbitrary and does not necessarily represent an underlying biological structure or mystical belief in the powers of the number three, although the grouping and sizes of the groups was a conscious decision.

The LF&G approach might have a positive impact on the development of easy-to-use software and hardware products. To determine the actual benefits of LF&G, members of the professional usability and design communities will need to learn how to use the principles and the forms created for practicing LF&G design and evaluation. At this point, LF&G has the potential, subject to appropriate research and practice, to develop into an empirically validated approach to define, describe, design, and predict the usability of human–computer systems.

9. CASE STUDIES OF REAL WORLD USE OF LF&G

9.1. Software Installation Interface Design and Evaluation

The initial use of the LF&G approach involved a port of an installation interface from one operating system platform to another. In this case, heuristic evaluation was used to refine an existing prototype design. Two LF&G expert heuristic evaluations × 2 usability professionals were conducted, the first at the end of unit testing prior to system verification testing, and the second prior to early ship (beta) program entry. A 21-item evaluation form based on a one to five rating scale with a separate rating for each principle and heuristic was created. Ratings for LF&G were arrived at by averaging the scores for the 7 items in each category. The overall ease-of-use rating was calculated as the mean of the three averages. These results, shown in Figure 1, were encouraging.

Further refinement of LF&G was deemed worth the effort. Based on feedback and analysis of the results from this study, some minor changes were made to the evaluation form, including the addition of "not applicable" as an answer option. This was deemed necessary to handle items that might be inappropriate within the evaluation context, for instance, the lack of online help in a wizard-like installation interface.

9.2. Hardware Product Out-of-Box-Experience (OOBE) Design and Evaluations

Additional development of the LF&G evaluation forms and methods has occurred within the context of customer installable network server hardware products. Modification of the evaluation and rating items to provide equivalent context-specific terminology for the OOBE evaluation was an initial adjustment. Debriefing of usability evaluation subjects revealed that participants had difficulty transferring the user interface concepts to the narrower hardware context, and a rephrasing into more concrete context-specific statements was needed to obtain a full set of responses. The OOBE evaluation results are listed in Table 3.

Severity ratings for identified problems were defined in a three-level hierarchy of severe, moderate, and minor. This provides nine potential levels for prioritization for problem resolution with severity as the first sort and LF&G category as the second sort. An example of a set of problems found under the lead category for one OOBE evaluation is shown in Table 4. In practice, additional columns for status and comments are added to the problem report and are used for tracking development and manufacturing progress in multidisciplinary team meetings.

During the design of the OOBE, six competitor evaluations of comparable network server products were conducted and rated using the LF&G approach. Products evalu-

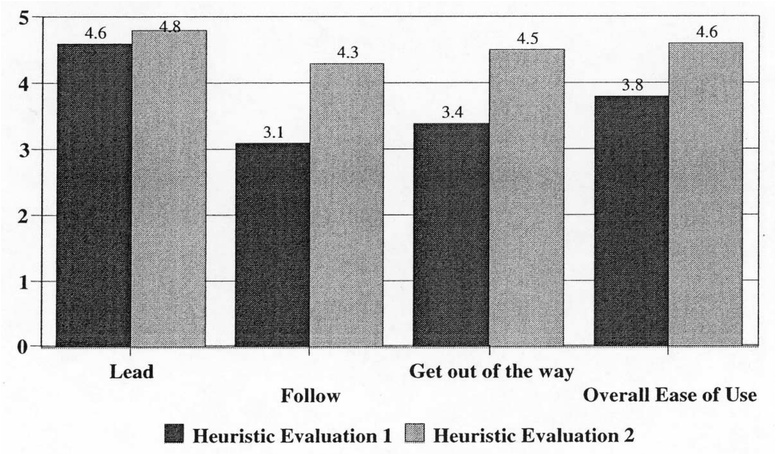

FIGURE 1 Software interface design and evaluation results using lead, follow, and get out of the way (LF&G). The LF&G heuristics were used to guide design recommendations and to rate ease of use during development, resulting in measurable improvement.

ated included internal and external competitive products commercially available at the time. Results did not indicate a bias for or against internal or external products.

The same expert evaluator was used throughout each of the competitor evaluations. Due to time and resource constraints, the evaluations were conducted using a combined task and cognitive walkthrough by the expert evaluator, observed by several human factors and hardware development professionals. Following the walkthrough, the expert and the observers completed LF&G evaluation forms. All ratings were combined for the overall ratings. These evaluation results are shown in Figure 2. The evaluation form was highly intuitive and self-explanatory, even to professionals outside the human factors discipline.

Following the competitive evaluations, both positive observations of compliance with LF&G principles and heuristics and observed problems due to violations of LF&G principles were compiled and listed as design points for the OOBE design of a customer setup network server product. These design points were tracked during development to ensure inclusion in the overall shipping and packaging design. A multidisciplinary UCD team lead by UCD and human factors professionals in-

Table 3: The Lead, Follow, and Get Out of the Way Principles and Supporting Heuristics—Out-of-Box Experience Evaluation Form Version

1. The way the components are shipped and unpacked leads the user to successful achievement of goals.
    - 1.1. How to unpack and setup the hardware is obvious and accessible.
    - 1.2. Errors are prevented by hiding or disabling potentially destructive actions.
    - 1.3. Components are labeled and the names used are distinct from each other and unambiguous in meaning.
    - 1.4. Instructions and physical affordances such as arrows, numbered labels, and so forth are clear, concise, and use familiar terminology.
    - 1.5. Where possible, components are preassembled, minimizing the possibility of error in assembly.
    - 1.6. The packaging and physical layout supports the natural task flow of unpacking and setting up.
2. Appropriate materials are obviously visible and readily available at the end of each step in the unpack and setup process.
    - 2.1. Components provide clear and obvious feedback whether physical, audible, or visible, of the completion of actions.
    - 2.2. Longer or more complex steps are broken down into shorter subunits to help indicate progress towards completion.
    - 2.3. Typical errors are anticipated and corrective actions are identified and described in case errors occur.
    - 2.4. Each step in the process has a clear beginning and a clear ending.
    - 2.5. Provide templates with preset measurements so that individual measurements do not need to be taken.
    - 2.6. Instructions and documentation address user tasks and goals.
3. The packaging and documentation get out of the way and do not interfere with efficient and effective unpacking and setup of the product.
    - 3.1. The number of individual actions, efforts, and tasks needed to complete the task of unpacking and setting up are minimized.
    - 3.2. Hardware and software components follow appropriate conventions and standards.
    - 3.3. The person setting up the hardware components is in control throughout the unpack and setup process; any step can be undone, redone, paused at, or exited from without difficulty.
    - 3.4. The packaging, documentation, and hardware are uncluttered with irrelevant details.
    - 3.5. The packaging of the components and the documentation support immediate productive use by a novice, but do not restrict someone with skill or experience from becoming more efficient.
    - 3.6. Shortcuts for performing tasks more swiftly are available or made obvious.

cluded active members from marketing, service planning, industrial design, information development, and shipping and packaging.

As can be seen in Figure 2, the results of prerelease prototype evaluations showed improvements in ratings over all available evaluated competitors. Having included both positive and negative findings from the competitor evaluations into the design points, this was to be expected. Prototypes at the equivalent of a software alpha release achieved ratings equal or better than generally available competitors. Pre-GA models (the equivalent of a software beta release) were rated above four points on the one to five point scale. At the time of this writing, external data to validate the internal experts' ratings had not yet been collected.

10. CURRENT STATUS OF LF&G

As can be seen from the case studies provided, LF&G is still in the early stages of development. On the one hand, it can be seen as a specific "best practices" approach to heuristic evaluation and user interface design practiced by a handful of practitioners of the art of usability. On the other, it can be seen as a theoretical model, at this point still pending empirical evidence to validate its claims of unifying principles for design and evaluation. It does not replace summative usability evaluation with end users in a laboratory setting, but it may provide additional design guidance and support at low cost during the development of hardware and software products. In the hands of experienced practitioners, it can be a useful tool. If nothing else, it attempts to synthesize much of the re-

Table 4: Reporting Severity of Problems Identified in Hardware Out-of-Box Experience Evaluation.

Severity	*Problem*	*Recommendation*
Severe	Difficulty lifting system	Add handles to sides of the system
Severe	Inner boxes not labeled and one was opened upside down	Label with contents and up and down directions
Severe	Product was not weight labeled for lifting requirements	Label with weight of product and recommendation for number of people required to lift. Place labels on both sides of the system
Moderate	Instructions for unpacking box located on top; therefore, obstructed once box is opened	Place unpacking instructions inside the box
Moderate	Drawn instructions did not match action needed for opening box (tabs)	Include correct drawings
Moderate	Instructions for installation not easily or quickly located; instructions discovered 55 min into procedure	Instruction sheet should be on top when box is opened, labeled, "how to proceed"
Minor	Lack of labeling to identify product	Label
Minor	Inner boxes were not labeled in the order to be opened	Label boxes in the order to be opened

Note. Severe problems are those that prevent completion of a task or have the potential to cause damage, loss of data, or physical harm to the user. Moderate problems are those that cause delays, confusion, or misdirection, but do not result in inability to complete the task. Minor problems do not cause significant delay but negatively impact the user satisfaction with the interaction experience.

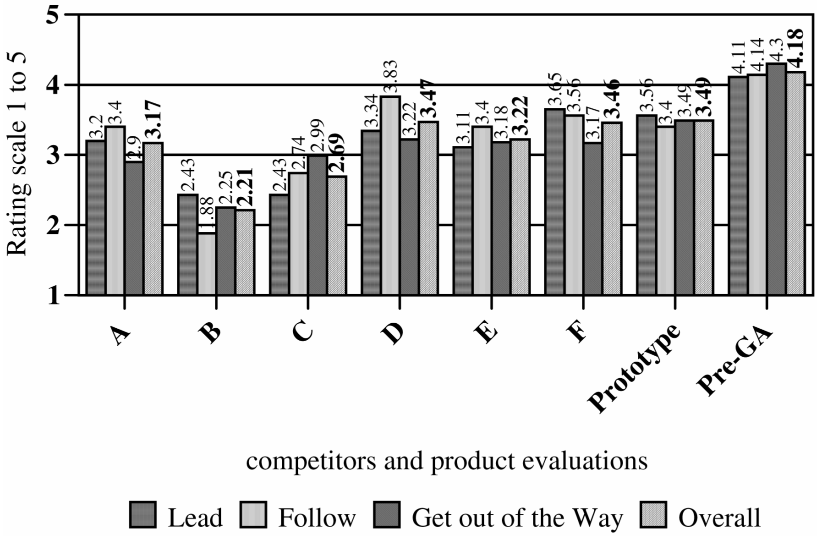

FIGURE 2 Out-of-box experience evaluations. Competitors versus internal prototypes.

search into heuristic evaluations into a framework that might be used to engender further research that could build a bridge between heuristics and observed results.

11. FUTURE DIRECTIONS

The LF&G approach has grown out of the practice of usability engineering, UCD, and the realities of software and hardware development. The initial applications of the LF&G approach to software and hardware development projects (during a period in which the technology industry experienced a contracting economy) have provided encouraging results that support the intended benefits and indicate that further research, study, and development of this approach is a worthwhile endeavor. Given the high face validity of the heuristics, it is not unreasonable to expect further positive results in the theoretical and practical development of the LF&G approach to designing and evaluating ease of use.

In an ideal world, the future directions for continued development of the LF&G approach might include the following activities:

- Use the LF&G heuristics and rating forms in expert heuristic evaluations across a variety of evaluators, locales, products, and knowledge domains to develop reliability estimates.
- Compare and contrast LF&G evaluations and ratings with other sets of heuristics and traditional usability evaluation metrics such as task completion, time on task, and user satisfaction.
- Extend the implementation of the LF&G rating and evaluation forms to Web-based surveys to support the ease of use of implementing LF&G as a usability evaluation rating method for end-user evaluations in addition to traditional metrics.

LEAD: The interface leads the user to successful achievement of goals.

Scale: 1=strongly disagree 2=disagree 3=neutral 4=agree 5=strongly agree	
1. Interface functions are obvious and available to the user.	1 2 3 4 5 NA
2. User errors are prevented by hiding, disabling, or confirming inactive or potentially destructive actions.	1 2 3 4 5 NA
3. Labels and names are distinct from one another - meanings are clear and unambiguous.	1 2 3 4 5 NA
4. Prompts are clear, concise, and use familiar terminology and language .	1 2 3 4 5 NA
5. Safe defaults are provided for user inputs - user has only to recognize information, not recall details..	1 2 3 4 5 NA
6. The natural work flow or task flows of the user are supported by and followed by the interaction with the interface.	1 2 3 4 5 NA
7. Overall, the interface leads the user to successful achievement of goals.	1 2 3 4 5 NA

LEAD: The interface leads the user to successful achievement of goals.

Scale: 1=strongly disagree 2=disagree 3=neutral 4=agree 5=strongly agree	
1. How to unpack and setup the hardware is obvious and accessible.	1 2 3 4 5 NA
2. Errors are prevented by hiding or disabling potentially destructive actions. Example: A label with a warning covers a switch, screw, etc. that would be damaging if loosened or tightened out of sequence.	1 2 3 4 5 NA
3. Components are labeled and the names used are distinct from each other and unambiguous in meaning.	1 2 3 4 5 NA
4. Instructions and physical affordances such as arrows, numbered labels, etc. are clear concise, and use familiar terminology.	1 2 3 4 5 NA
5. Where possible, components are pre-assembled, minimizing the possibility of error in assembly..	1 2 3 4 5 NA
6. The packaging and physical layout supports the natural task flow of unpacking and setting up.	1 2 3 4 5 NA
7. Overall, the way the components are shipped and unpacked leads the user to successful achievement of goals.	1 2 3 4 5 NA

FIGURE 3 Examples of the LEAD section for user interface and hardware out-of-box experience (OOBE) rating evaluation forms. Future versions of the rating evaluation forms may include additional points on the scale.

- Develop variations in addition to software interface and hardware OOBE evaluation forms for use in different knowledge domains if needed (e.g., online help systems, hard copy documentation, Web interaction design, mobile computing devices and interfaces, pervasive computing, or embedded technology interfaces).
- Refine the LF&G rating forms to improve the robustness of the instrument itself. This may include increasing the rating scale from a 5-point scale to a 10-point scale or randomizing the order in which the rating questions appear, or other changes that professionals in the measurement and evaluation field may deem advisable. A sample of the Lead section of the two current forms is provided in Figure 3 for those interested in comparisons between the heuristics and the evaluation forms.
- Use the LF&G rating forms in conjunction with user testing to determine if correlation between expert heuristic and end-user ratings of the same interface are strong enough to be useful to business decision makers.
- Develop standard training materials and protocols to help control for variance in individual evaluators' understanding, interpretation, and application of the LF&G heuristics.
- Incorporate LF&G into automated tools to ensure the quality and consistency of UCD and evaluation work across of variety of contexts, users, and products.
- Elaborate on the research base that supports the principles and heuristics to support training and education in the design and evaluation of usability.

The world in which LF&G has been and continues to be developed is not ideal, however, and the extent to which the aforementioned vision will be made manifest is yet to be determined.

REFERENCES

Borges, J. A., Morales, I., & Rodriguez, N. J. (1996). Guidelines for designing usable World Wide Web pages. In *Proceedings of the CHI '96 Conference Companion on Human Factors in Computing Systems: Common Ground* (pp. 277–278). New York: ACM Press.

Butler, K. A. (1996). Usability engineering turns 10. *Interactions, 3,* 59–75.

Doubleday, A., Ryan, M., Springett, M., & Sutcliffe, A. (1997). A comparison of usability techniques for evaluating design. Symposium on Designing Interactive System. In *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (pp. 101–110). New York: ACM Press.

IBM Ease of Use. (n.d.). *Web design guidelines—Print view.* Retrieved January 7, 2002, from http://www-3.ibm.com/ibm/easy/eou_ext.nsf/Publish/572PrintView

Jacobsen, N. E., Hertzum, M., & John, B. E. (1998). The evaluator effect in usability tests. In *Proceedings of the CHI '98 summary: Human factors in computing systems* (pp. 255–256). New York: ACM Press.

Jeffries, R., Miller, J. R., Wharton, C., & Uyeda, K. M. (1991). User interface evaluation in the real world: A comparison of four techniques. In *Human Factors in Computing Systems Conference Proceedings on Reaching Through Technology* (pp. 119–124). New York: ACM Press.

Kamper, R. (2001). *Lead, follow and get out of the way: A practical approach to ease of use design, evaluation, and achievement* (Tech. Rep. No. 51.999). Austin, TX: IBM.

Kantner, L., & Rosenbaum, S. (1997). Usability studies of WWW sites: Heuristic evaluation vs. laboratory testing. In *Proceedings of the 15th Annual International Conference on Computer Documentation* (pp. 153–160). New York: ACM Press.

Köykkä, M., Ollikainen, R., Ranta-aho, M., Milszus, W., Wasseroth, S., & Friedrich, M. (1999, November). *Usability heuristic guidelines for 3D multiuser worlds.* OzCHI '99 Conference, Wagga Wagga, New South Wales, Australia. Retrieved January 8, 2002, from http://www.csu.edu.au/OZHI99/program.html

Lewis, J. R. (2002). Psychometric evaluation of the PSSUQ using data from five years of usability studies. *International Journal of Human–Computer Interaction, 14,* 463–488.

Muller, M. J., Matheson, L., Page, C., & Gallup, R. (1998). Participatory heuristic evaluation. *Interactions, 5,* 13–18.

Nielsen, J. (1992). Finding usability problems through heuristic evaluation. In *Proceedings of CHI '92* (pp. 373–380). New York: ACM Press.

Nielsen, J. (1993). A mathematical model of the finding of usability problems. In *Proceedings of INTERCHI '93* (pp. 206–213). New York: ACM Press.

Nielsen, J. (1994). Enhancing the explanatory power of usability heuristics. In *Proceedings of ACM CHI '94* (pp. 152–158). New York: ACM Press.

Nielsen, J., & Molich, R. (1990). Heuristic evaluation of user interfaces. In *Proceedings of CHI '90* (pp. 249–256). New York: ACM Press.

Nielsen, J., & Phillips, V. L. (1993). Estimating the relative usability of two interfaces: Heuristic, formal, and empirical methods compared. In *Proceedings of INTERCHI '93* (pp. 206–213). New York: ACM Press.

Polson, P. G., & Lewis, C. H. (1990). Theory-based design for easily learned interfaces. *Human-Computer Interaction, 5*, 191–220.

Savage, P. (1996). User interface evaluation in an iterative design process: A comparison of three techniques. In *Proceedings of CHI '96* (pp. 307–308). New York: ACM Press.

Sears, A. (1997). Heuristic walkthroughs: Finding the problems without the noise. *International Journal of Human–Computer Interaction, 9*, 213–234.