**FINAL EXAM, IAT432, SPRING TERM 2013 ANSWER KEY**

**Instructions:** The scope of this exam covers the entirety of the course. You will have about 3 hours to complete your work; this should be plenty of time. During the exam, you may ask questions of the invigilator, BUT YOU MAY NOT EXCHANGE INFORMATION WITH ANOTHER STUDENT. Cellphones and other potential memory aides must turned off and packed away out-of-sight. The same is true of other electronics such as computers. The test has the following components. Plan your time accordingly:

| | | |
|---|---|---|
| 20 | Multiple Choice (2.5 pts each) | 50 pts |
| 10 | Short Answers (2 pts each) | 20 pts |
| 1 | Long essay (choice from 2) | 30 pts for one complete answer |

Note that there are two long essay questions *BUT ONLY ANSWER <u>ONE</u> OF THEM – YOUR CHOICE*. There are no extra credit items on the test.

If you have not enough paper for your answers, please ask the invigilator for more paper. Remember to put your name, student number, and question number on any continuation pages and on each page of this exam.

**I. Multiple Choice Items:** Circle the letter of your choice in each of the next items. Just one alternative is correct UNLESS stated otherwise.

1. According to Dix in his chapter on Evaluation Techniques (Ch. 9) a cognitive walkthrough as a usability tool refers to . . . .

   A. *<u>a "hands-on" review of steps as choices, thoughts a design requires to complete an action, mimicking what a real user might experience..</u>*
   B. a rewrite of interface design steps by an expert user, mimicking words a user might have to describe an action.
   C. a test of web interfaces by a large sample of randomly chosen people, looking for the errors and mistakes they make.
   D. a computer game where deliberately misleading directions are given to test subjects finding a prize in a virtual building. E.
   E. all of the above.

2. In software design, "wizards" are . . . .

   A. *<u>where complex actions are broken up into step-by-step instructions, guiding novice users to a desired outcome.</u>*
   B. experts who offer "how to" advice on blogs or bulletin boards (BBS).
   C. error messages that identify machine code and command faults for programmers.
   D. what Apple Store employees are called at the stores' "genius bar".
   E. usability testing firms that refine and improve faulty computer software and instructions.

3. Nielson's "thinking aloud" system is used to . . .

    A.  get expert opinion on need for new software products, also called a "focus group".
    ***B.  get users to speak aloud problems and blocks while trying out new software***
    C.  count the number of negative comments as potential buyers look over new products
    D.  as an exercise where computer users speculate aloud about their perfect "dream machine".
    E.  follow consumers through a store recording the products they pick-up and test out what people will pay for certain software and hardware features

4. The following item is essential to scientific method and research . . .

    A.  institutional permission; especially if humans are involved
    B.  a sample size numbering at least 30 subjects
    ***C.  a null hypothesis***
    D.  inferential statistics
    E.  machine readable data

5. An hypothesis is. . . .

    ***A.  testable question suggesting a possible relationship***
    B.  is a confounding set of variables
    C.  only for quantitative data
    D.  for laboratory research only
    E.  an explanation for failed results.

6. Reliability tests . . . .

    ***A.  the time-after-time consistency or stability of a research result or outcome***
    B.  the correspondence or fit of the measure to the concept or construct measured
    C.  is also known as the Hawthorne test
    D.  must always be done by a second group of subjects or researchers
    E.  is ethically required in all human subject studies

7. In a statistically "normal" distribution how much of population outcomes are between plus and minus one (1) standard deviation . . .

    ***A.  about 2/3rdsd or 68%***
    B.  about a third or 33%
    C.  about 6%
    D.  can't be estimated; varies greatly
    E.  always 98%

8. Which statement is TRUE:

   **A.** ***You may have more than one MODE but only one MEDIAN in a given distribution***
   B. There is only one MEDIAN and one MODE in a distribution
   C. Both MEDIAN and MODE are the same as the MEAN
   D. When the **mean, media and mode** are the same value, the distribution is skewed
   E. Mean, media and mode are measures of **variance**.

9. Surveys have the following advantages and disadvantages (**circle ALL that apply**):

   **A.** ***Can be representative through random selection of subjects***
   **B.** ***Are often costly and slow***
   **C.** ***Have high refusal rates***
   **D.** ***Often have difficult coding and analysis.***
   E. Need statistical knowledge to be really effective

10. "Field" research is . . .

    A. Mainly done in large, open agricultural areas.
    **B.** ***Done where people live and work***
    C. Typically done in controlled laboratories
    D. Uses only secondary or historical data
    E. Especially fast and cheap to complete

11. User Experience (abbreviated "UX") is . . .

    A. Specific just to the users speed in learning a designed interface
    B. Focuses on feelings or "felt" experiences in using an interface
    **C.** ***Widely covers issues of performance, affect and user satisfaction***
    D. Only directly observed through user behaviors
    E. Always fully "objective" or non-judgmental

12. A "cultural interface" . . . .

    A. uses informal language to suit users (like teens)
    B. uses art forms, aesthetics, stylistic references that are meaningful and attractive to the target user
    C. uses suitable metaphors for commands that mimic or echo other familiar
    **D.** ***all of these (A, B, C and E)***
    E. Employs "hybridity"

13. The idea of "hybridity" in an interface . . . .

    A.  Suggests what the interface is used for, to do.
    ***B.  design bridge between the control interface and user culture***
    C.  Builds--in "surprise" elements to entertain users.
    D.  A design that successfully blends two distinct cultures
    E.  All of above.

14. In the 1980s, computers developed WIMP interfaces, which are with us today. The term means:

    A.  windows, interactivity, mouse, picture
    B.  wi-fi, icons, microcomputer, printers
    ***C.  windows, icons, mouse, pointer***
    D.  web, internet, mouse, printer
    E.  web, internet, microprocessor, picture

15. An example of a metaphoric representation in interface design is . . .

    A.  Combining functional and cultural interfaces
    B.  Using a "genre" like a family album
    ***C.  Using a virtual bookshelf to show on-line books bought***
    D.  Using fixed standards to make use easier
    E.  A paradigm shift, say from a mouse-based to a touch-based interface

16. Techniques of testing usability, such as "heuristic design" and other methods leave aside one very important consideration:

    A.  Cost
    B.  Multimedia
    C.  Culture
    ***D.  Aesthetics***
    E.  Language

17. Apple products are famous for . . . (choose best answer)

    ***A.  Design aesthetics***
    B.  Providing compact sizes for products
    C.  Dispensing with floppy disk drives and other old tech
    D.  Low costs
    E.  Availability world--wide

18. CHI usability trends for the last decade show amongst technology manufacturers . . . . .

 A. Increased usability testing
 ***B. Increased usability testing that is quantitative (best answer, Partial credit for A above)***
 C. Increased use of informal analytics, such as heuristic evaluation.
 D. Steady use of outside consultants to test usability
 E. Better response to user suggestions and complaints

19. As a sample size is increased (assuming random selection) . . .

 A. Costs of the survey go down
 B. It takes less time
 ***C. There's less chance of error from a few extreme (outlier) cases***
 D. There is more social excitement in the community
 E. All of the above

20. The T-Test (also known as Gauss' students T) . . . .

 ***A. checks the likelihood that two sampled means are really different beyond chance***
 B. checks mainly the consistency of a common hot beverage
 C. is a descriptive statistic
 D. is a correlation coefficient
 E. is really none of the above

**SHORT ANSWER**: You should be able to answer these items in a few sentences or short paragraph.

21. Explain what an **heuristic evaluation** of usability tries to do? Is it a complete test or just a good way to roughly gauge the quality of design?

**Heuristic evaluation uses a small group of experts in usability (nominally about 4-5) to try out software and systems whilst carefully noting issues and illogic, conversely too, good design and flow. Usually they are new to the software so that first impressions are gained – and from their expertise in-depth commentary and suggestions for solutions. This method caches roughly 80% of errors and thus is fast, inexpensive and suggestive of fixes to new product design.**

22. Nielsen outlines 10 Usability Heuristics. Discuss one that you consider most important, giving reasons for your choice.

**Neilson's 10 "heuristics" and briefly why important:**

**Summary:** The 10 most general principles for interaction design. They are called "heuristics" because they are more in the nature of rules of thumb than specific usability guidelines.

**Visibility of system status**

The system should always keep users informed about what is going on, through appropriate feedback within reasonable time.

**Match between system and the real world**

The system should speak the users' language, with words, phrases and concepts familiar to the user, rather than system-oriented terms. Follow real-world conventions, making information appear in a natural and logical order.

**User control and freedom**

Users often choose system functions by mistake and will need a clearly marked "emergency exit" to leave the unwanted state without having to go through an extended dialogue. Support undo and redo.

**Consistency and standards**

Users should not have to wonder whether different words, situations, or actions mean the same thing. Follow platform conventions.

**Error prevention**

Even better than good error messages is a careful design which prevents a problem from occurring in the first place. Either eliminate error-prone conditions or check for them and present users with a confirmation option before they commit to the action.

**Recognition rather than recall**

Minimize the user's memory load by making objects, actions, and options visible. The user should not have to remember information from one part of the dialogue to another. Instructions for use of the system should be visible or easily retrievable whenever appropriate.

**Flexibility and efficiency of use**

Accelerators -- unseen by the novice user -- may often speed up the interaction for the expert user such that the system can cater to both inexperienced and experienced users. Allow users to tailor frequent actions.

**Aesthetic and minimalist design**

Dialogues should not contain information which is irrelevant or rarely needed. Every extra unit of information in a dialogue competes with the relevant units of information and diminishes their relative visibility.

**Help users recognize, diagnose, and recover from errors**

Error messages should be expressed in plain language (no codes), precisely indicate the problem, and constructively suggest a solution.

**Help and documentation**

Even though it is better if the system can be used without documentation, it may be necessary to provide help and documentation. Any such information should be easy to search, focused on the user's task, list concrete steps to be carried out, and not be too large.

23. We have covered four (4) methods of examining usability. Each has virtues and faults. Contrast any two, finding one advantage and one disadvantage to each.

**Heuristic Evaluation: Fast, cheap, expert, has depth captures about 80% problems. But not representative, doesn't use actual novice, customers, may suffer because "experts" familiar with product.**

**Usability Study, such as "thinking out loud" with a small novice user sample. Gets hangups in software and hardware design, use flow, poor documentation, awkward controls. But again is not necessarily a large and representative sample. Testing takes place in lab, so may be unlike home or work use.**

**Controlled Experiment: Gets at core validity and reliability of what makes the software/hardware/design work well, poorly.  Allows high control over multiple independent variables. Allows tight control over intervening variables. But because of lab setting external validity (how alike at home/work use is to lab) can be low. Expensive to do, especially large scale – i.e. many subjects tested.**

**Cued-recall debrief: Has characteristics (pro and con) of usability, but builds in an aided recall of first time use with each subject. Subjects can respond in depth to feelings, problems, good features and any confusion. Time consuming and expensive if a lot of subjects.**

24. In a "depth" interview, the questioner really is seeking what kind of information?

**Sensitive and contingent items relating to life-style, emotion, goals, confusion, annoyances, conformity with expectations from the design. In other words, an elaboration of likes and dislikes but backed by prior expectations and experience. Many of these outcomes are the same as "heuristic evaluation".**

25. In a questionnaire, what kind of items should be left for last in the interview?

**Leave for last those items that may anger or impinge on the respondents' sense of privacy or trust; items like age, income, health problems, personal relationships, job satisfaction, political votes and beliefs. Religion. These items can evoke a negative reaction and the subject terminating the interview. By leaving until last you will have (a) built trust, rapport through tame questions, and (b) if interview prematurely ends, you have most of the data anyway.**

26. Give an example of a Likert questionnaire item. What is a useful feature of this type of question?

**A Likert item is simply a statement which the respondent is asked to evaluate according to any kind of subjective or objective criteria; generally the level of agreement or disagreement is measured. It is considered symmetric or "balanced" because there are equal amounts of positive and negative positions.[6] Often five ordered response levels are used.**

**The format of a typical five-level Likert item, for example, could be:**

1.   **Strongly disagree**
2.   **Disagree**
3.   **Neither agree nor disagree**
4.   **Agree**
5.   **Strongly agree**

27. Contrast a structured v. an unstructured interview. What is one positive and one negative feature of each type? Give an example of a characteristic question from each type.

**Structured: Responded replies to scales and pre-worded checklists. Answer alternatives are put in categories and words of the researchers. In this sense, the interview is precoded and the responded places his or her answer in the appropriate category. For example, see Likert items above. Or a multiple choice exam.**

**Unstructured: Respondents reply in their own words, which are recorded close to verbatim by the interviewer. Often tape recording used. Later, responses have to be coded, categorized and counted. Examples: response in one's own words; an essay exam.**

28. How can one assess or measure validity. Describe at least one way (there are several).

**Validity of a question is the degree to which it measures what it is supposed to measure in the view of the researchers and research community. This is not the same as reliability, which is the extent to which a measurement gives results that are consistent. Within validity, the measurement does not always have to be similar, as it does in reliability.**

**Construct validity: do judges see the links between measurement operations (questions, say) and the concept or idea being represented.**

**Criterion validity: does the measure predict or identify the end-state or criterion of the variable. The *mortal* injury results in death.**

**Concurrent. Are there two conceptually distinct measures of a concept – and do they agree across respondents, i.e. high humidity with high temperature predicting environmental comfort.**

29. What is a "confounding" or intervening variable? How does one control them?

**As an example, suppose that there is a statistical relationship between ice cream consumption and number of drowning deaths for a given period. These two variables have a positive correlation with each other. An evaluator might attempt to explain this correlation by inferring a causal relationship between the two variables (either that ice cream causes drowning, or that drowning causes ice cream consumption). However, a more likely explanation is that the relationship between ice cream consumption and drowning is spurious and that a third, confounding, variable (the season) influences both variables: during the summer, warmer temperatures lead to increased ice cream consumption as well as more people swimming and thus more drowning deaths.**

**Generally, confounding variables confuse or cloak the true causes of phenomena. In designing studies, one tries to identify or eliminate these defects, confusions. Laboratories are specially constructed spaces that try to isolate away these influences (such as construction noises during hearing acuity testing).**

30. Briefly describe "cued recall debrief". Why does this procedure work well for assessing usability? Describe one big advantage.

**See item #23, last method listed.**

**Extended Essay Questions:** **Please choose one of the following two (2) to answer**

31. Controlled Experiments: I want to conduct a controlled experiment to test whether the interaction style found Interface A is better than the style found in Interface B for finding and clicking on a target in an application list for Mac OS X.

**The point of this item is to create (a) formal comparison (the two interfaces below), that (b) shows the influence of one or more independent variables on an (c) outcome or dependent variables with (d) adequate controls to avoid confounding or spurious results**

**Interface A**



*When you mouseover entries in the application list, each stays the same size.*

Answer the following . . .

31A . What hypothesis would you recommend for this experiment?

**Does the mouseover raising the size of icons speed use and reduce errors**

31D. What would you choose for independent variables? **See 31A**

**Interface B**



*Mouseover enlarges entries*

31B. What would you recommend as your null hypothesis?

**There is no improvement on speed or error rate due to mouseover size change**

31C. What are your dependent variables?

**In this example, speed and errors in completing a multi-program selection and use exercise (say word processing with photoshop work on images and email of results to dropbox).**

32A.  Below are two instructions that have the same result. The right-hand one is probably more usable. Why?

- **Changes progressively as one moves through a complex multistep activity.**
- **Has contrasting graphics that has a 'progress bar" Older one has no indication of progress or why/where a step-by-step chain stops.**

32B. Discuss advances you see in user interface aesthetics and functionality that are more a result of graphics, layout, instructions -- features such as treating user guidance as interconnected parts of a functional machine able to do complex things.

- **This is a "wizard" that confirms each user action with a graphically mapped system response. It is reassuring to the novice and gives diagnostic support to the expert**

32C. What recent advances, features do you see in user design that seem especially clever and helpful, such as found in tablet and smartphone systems. Describe and justify at least two.

- **Touch interface**
- **Icons rather than mouse moves or command lines; better than WIMP**
- **Interconnected steps such as asking if one wants to turn on wi-fi when connectivity is needed.**
- **Design conformity among phones, tables and desktop/laptops. Ex. Windows 8 family,  OSX, iOS for Mac.**

32D. It has been said that computers, smartphones and tablets are in a design race for dominance. Increasingly, the Apple desk machines resemble the agility of the iPad and iPhone software.  Windows has aligned phone, tablets and its new PC operating system (windows 8) to look much alike.  Discuss design, functionality, uniformity and portability for user interface design. Use examples.

- **Transfer of training and orientation infamilies of device**
- **Using physical actions, metaphors– pinches, wipes, etc. to alter images and actions**
- **Multisourcing telecoms – wi-fi, network, hotspots, VPN with memory hoding access keys**
- **Transparent use of "cloud" assistance for backup, complex processing such as for the Siri assistant and voice-to-text conversion.**
- **Integration with other devices such as motor cars' "infotainment", new generation of TVs and home audio systems or multimedia centres.**

text