Defending Design Decisions with Usability Evidence: A Case Study

Erin Friess
Carnegie Mellon University
English Department
5000 Forbes Avenue
Pittsburgh, PA 15213 USA
efriess@andrew.cmu.edu

Abstract
This case study takes a close look at what novice designers discursively use as evidence to support design decisions. User-centered design has suggested that all design decisions should be made with the concern for the user at the forefront, and, ideally, this concern should be represented by findings discovered within user-centered research. However, the data from a 12-month longitudinal study suggests that although these novice designers are well versed with user-centered design theory, in practice they routinely do not use user-centered research findings to defend their design decisions. Instead these novice designers use less definitive and more designer-centered forms of evidence. This move away from the user, though perhaps unintentional, may suggest that design pedagogy may need to be re-evaluated to ensure that novice designers continue to adhere to the implications of user-centered research throughout the design process.

CHI 2008, April 5–10, 2008, Florence, Italy.

ACM 978-1-60558-012-8/08/04.

Keywords
User-centered design, design pedagogy, discourse analysis, decision-making practices

ACM Classification Keywords
H.5.2. User Interfaces: Theory and Methods

Introduction
Today, the concept of user-centered design (UCD) has been "entirely accepted" by the field of human-computer interaction (HCI) [11]. While there is some disagreement as to what constitutes UCD [8,9], a key component of all forms of UCD is user involvement. The most heralded benefit of this user involvement in UCD is that by incorporating users throughout the process companies will ultimately save money and reputations by solving potential problems during development rather than learning about shortcomings only after deployment [5].

If user-centered research and its subsequent findings are so important to the success of UCD, then it would seem to follow that designers would use user-centered research findings as evidence to support claims regarding design decisions. In other words, when designers are contemplating what design choice to make, reasons that include user-centered data, be it front-end analysis or usability-based findings, would seemingly be the gold standard. Therefore, in this 12-month case study, I observed the meetings of novice designers to determine what they used discursively as evidence to support their design decisions.

Method

The group I observed consisted of 26 designers in their natural workplace setting. These designers were hired by the United States Postal Service (USPS) to redesign several important paper-based documents for the USPS. This team was under the direction of a faculty principal investigator, and was led by a project manager and an assistant project manager, both with Master of Design degrees. The remaining members of the group were all Master of Design or PhD in Design graduate students. Despite composition changes at the onset of the semesters and summer break, the number of graduate students held steady at 18 throughout the course of the project. Save for the PI and the project manager, the majority of the students had no professional design experience and the only client-based design experiences were from academic courses. Those students who did have professional design experience (7 students) generally had less than two years of experience.

This team was familiar with concepts of user-centered design and the potential benefits of usability practices. The design department's mission statement includes: "We believe in learning by doing in a human-centered design process and in the value design brings to a range of human activities." Further, at some point during their first year of graduate studies, all of the students would have to take a course on research methods for user-centered design. Additionally, all new team members received a document created by the project manager entitled "Overview of Approaches and Methods for User Centered Research and Design." This document states: "Fundamental to our approach is that designers themselves conduct research as an integral part of the creative process" (personal correspondence). There was no separately delineated usability group; the designers did all research activities themselves. Furthermore, according to the statement of work, approximately 35% of the designers' time and budget was to be researching users and testing prototypes, far beyond the 19.25% of the total budget in typical design projects [12].

The general work process for the group involved working in teams of 3-5 students for two weeks on projects assigned to them by the project manager. These assignments included prototyping, editing, arranging information, programming, developing new graphics, preparing for client meetings, or conducting user-centered research, among other tasks. The user-centered research conducted by the designers included field research, interviews, think-aloud walkthroughs, focus groups, questionnaires, and card-sorting exercises with both expert and novice users. After two weeks, the entire group would meet together. The general goal of these meetings was to learn what the other groups were doing, to come to a consensus on a way forward, and to decide on the tasks to be

accomplished for the next meeting. My participation with the group was limited to recording the meetings. After one year, I had collected 10 usable meetings, lasting between 30 minutes and 2.5 hours.

Discourse Analysis and Evidence

Using discourse analysis techniques, I transcribed all the spoken exchanges that occurred within the meetings. Discourse analysis "can provide rigor and systematicity that is sometimes missing in the critical tradition" while also allowing qualitative interpretations that are more difficult in more quantitative endeavors, such as prototype comparison [7]. In order to determine how the designers used the research within their decision-making behaviors, I coded each transcript for the basic argument elements of claims and grounds (which is also called evidence or support) as established in Toulmin's terminology [16]. Toulmin, a leading figure in the practical (rather than formal) argumentation movement of the 20th century, defines a claim to be a statement that includes "a standpoint regarding a subject," and grounds/evidence to be "certain facts on which the claim is based" [17]. Examples of claims from the meetings include:

- "We don't need mail prep anymore."
- "I think that it is presumptuous to think that at this time."
- "So, what if these people do come across multiple citations, wouldn't that, uh, just confuse them?"

In formal logic, claims such as these would have to be supported with some kind of support in the form of grounds/evidence [17]. However, as can be expected in this everyday argument, many claims are made without any direct support.

In addition to claims, I coded for grounds, or support. Support for a claim can be given without solicitation or may be requested by others, but the evidence must always support a claim. In other words, while a claim might be able to exist without evidence, evidence cannot exist without its claim counterpart.

For this analysis, I borrowed and slightly modified the support categories established by Deanna Kuhn in The *Skills of Argument* [10]. Kuhn conducted interviews with her subjects and found they tended to support their own claims in one of three ways: fact-based evidence, non-evidence, and pseudoevidence.

Fact-based Evidence

Fact-based evidence, evolving from Kuhn's notion of genuine evidence, is separate from the claim itself and includes at least an attempt by the participant to support the claim by pointing to a source of origin outside the speaker's self. For example: "We should use, I don't know, at least a 10/12 typeface…'cause remember that Bob, um, Bob didn't like that small typeface we used on the P2."

In this example, the speaker claims that "we should use…at least a 10/12 typeface." Without solicitation, the designer supports that claim with the evidence that "Bob [a usability study participant who did scenario-based testing] didn't like the small typeface we used on the P2." By pointing to something that actually happened during a usability study, the designer is employing fact-based evidence to support the claim. Fact-based evidence includes any support for a claim

that references a usability study, a written text, or a perceived authority figure.

Non-evidence
Non-evidence is "the effect as evidence of its cause," or any attempt to support a claim by simply repeating the claim" [10]. For example: "…we shouldn't use a small typeface because it's small."

In this example, the designer makes the claim that they shouldn't use a "small" typeface, and then goes on to support this claim simply by repeating the claim.

Pseudoevidence
Kuhn suggests that "pseudoevidence can…be thought of evidence by illustration" [10]. Further, "pseudoevidence takes the form of a scenario, or script, depicting how the phenomenon might occur" [ibid]. For coding purposes, pseudoevidence is any attempt by the designers to support a claim that cannot be coded as non-evidence and that does not point to some source outside themselves. For example: "I think it would be a mistake if we have a small typeface because it'll make it hard especially for the elderly to read it."

Here the designer claims that "it would be a mistake if we have a small typeface." The support of "because it'll make it hard especially for the elderly to read it" is not non-evidence because it does not simply repeat the claim. However, the support does not point to any research or authority figure outside of the designer's self that would support the claim. The support is a hypothetical story about one possible outcome (indeed, "it'll" posits the evidence for the current claim in the future). This hypothetical story originates within the speaker's self, and thus is pseudoevidence.

Extended Example
Perhaps it is best to compare these types of support, particularly fact-based evidence and pseudoevidence, with an extended example. In the example in Table 1, the group is discussing the general information architecture of their document. They had previously decided that there were three parts of the architecture: shape (such as letter or parcel), class (such as First-Class Mail or Media Mail), and service (such as Retail or Bulk).

Speaker	Turn	Thought Unit	Utterance
Linda	64	A	I know that um, that we thought that the order would, um, be best with shape first
		B	But after our, um, interviews this week, um, well the employees found it, um, found it kinda hard.
Ivan	65	A	So, um what are you suggesting
		B	That we start with class?
Linda	66	A	Umm
Tom	67	A	No, we can't start with class
		B	That just, um, doesn't, um that won't help the employees make a decision

Table 1: Extended Example of Fact-based Evidence and Pseudoevidence

Linda begins in 64A with the claim "we though that the order would, um, be best with shape," but then goes on to deny that claim in 64B with "but after out, um, interviews this week, um, well, the employees found it, um, found it kinda hard." This denial is fact-based evidence because it refers to actual usability testing conducted with USPS employees who found the architecture "kinda hard." After Ivan asks Linda if she is suggesting that they "start with class," Tom offers a claim in 67A with "we can't start with class." Tom supports his claim in 67B with "that just, um, doesn't, um, that won't help the employees make a decision." In this support, Tom doesn't refer to any usability testing nor to any other outside authority to help defend the claim "we can't start with class." Instead, Tom offers a hypothetical story that starting with class "won't help the employees make a decision," thereby making the support pseudoevidence. Had Tom said, "our previous usability studies have shown that starting with class doesn't help the employees make a decision," then the support would have been classified as fact-based evidence.

Coding

After an initial training session, three raters coded a 5-page sample set for the three types of support and subsequently made appropriate adjustments to the coding criteria. Two raters, working independently, coded identical samples of approximately 12% of the total data set. The coding scheme was tested for reliability using Cohen's Kappa, and yielded a kappa of 0.71, which is in line with the typical kappas report in group argument structure research [2, 13].

Results

Claims and Support

The designers made many claims without any support whatsoever. On average, only 57.8% of the claims made during the meeting were accompanied in some way by any kind of support, be it fact-based evidence, non-evidence, or pseudoevidence (Table 2). Further, the number of claims per hour ranged from 28.6 (meeting 8) to 48.6 (meeting 7), with an average of 34.6 claims per hour. The number of support per hour ranged from 10.0 (meeting 5) to 31.2 (meeting 1) with an average of 20.2.

Meeting	Length (in min)	Claims	Support	% Claims with Support
1	75	46	39	84.8
2	68	39	31	79.5
3	101	50	22	44.0
4	80	41	18	43.9
5	42	20	7	35.0
6	155	82	55	67.1
7	21	17	7	41.2
8	88	42	20	47.6
9	32	19	12	63.2
10	42	28	20	71.4
Avg.	*70.7*	*38.7*	*23.1*	*57.8*
Std. Dev.	*37.2*	*18.6*	*14.2*	*16.7*

Table 2:Claims and Support within Meetings

There does not appear to be any correlation between 1) the length of the meeting and the amount of claims, support, or percentage of claims with support, 2) the

meeting number and the number of claims, support, or percentage of claims with support, or 3) the simple number of claims and the number of claims with support.

Meeting	Fact-based Evidence	Non-evidence	Pseudo-evidence
1	6	4	29
2	7	1	23
3	5	0	17
4	8	0	10
5	1	0	6
6	20	2	33
7	2	0	5
8	5	0	15
9	4	0	8
10	7	0	13
Avg.	*6.5*	*0.7*	*15.9*
Std. Dev.	*5.0*	*1.3*	*9.2*

Table 3: Amount of the Three Types of Support Per Meeting

Types of Support

When looking at the support alone, pseudoevidence far surpasses the use of fact-based evidence and non-evidence combined. The results, summarized in Table 3, indicate that 70.5% of the time some kind of support is given for a claim, it will be deployed by the designer in the form of pseudoevidence. Further fact-based evidence (which includes all the times usability findings are used as evidence) is used only 27.6% of the time, while 3.2% of the time the designers used non-evidence. The amount of fact-based evidence per hour ranged from 1.4 (in meeting 5) to 7.5 (in meeting 9), with an average of 5.6 uses of fact-based evidence to support a claim per hour. The amount of non-evidence per hour ranged from 0 (in seven meetings) to 3.2 (in meeting 1), with an average of 0.5 uses of non-evidence to support a claim per hour. The amount of pseudoevidence per hour ranged from 8.6 (in meeting 5) to 23.2 (in meeting 1), with an average of 14.1 uses of pseudoevidence per hour. There does not appear to be any correlation between: 1) the length of the meeting and the number of fact-based, pseudoevidence, or non-evidence use, or 2) the meeting number and the amount of fact-based, pseudoevidence, and non-evidence.

Analysis

Clearly, the novice designers in this study gravitate toward the hypothetical pseudoevidence (which they use 70.5% of the time they support a claim) over the research-borne fact-based evidence (which they only use 27.6% of the time to support a claim). This reliance on pseudoevidence is rather surprising, given the fact that this group was mandated by their client to dedicate 35% of their time and budget to user research and usability studies, which would seemingly produce ready-made fact-based evidence. Instead of incorporating the findings of usability sessions into their talk when engaged in decision-making discussion, this group routinely relied on support with no definitive referent, be it a usability result or a reference to some other authority. By seemingly by-passing the findings from usability (as apparent in their lack of fact-based evidence) and offering instead self-generated pseudoevidence, these designers, who are well-accustomed with UCD and often espouse its benefits, are ostensibly only doing half the UCD process: they

are doing the user research, but they aren't using it to help drive decision-making.

This group's reliance on pseudoevidence, coupled with separate studies by Sugar and Dong that suggest novice designers do not use usability studies in manners that are efficient or effective for design, may imply that there is some obstacle for novice designers in learning how to maintain user-centered research findings throughout the design process [4, 15]. While our design students can conduct usability studies and can create products, how do we know that the final products are indeed receptive to the needs of the user outlined in usability study? Most guides to user-centered design and usability practice focus on, specifically, usability sessions and an iterative design process; however, the steps that explain how a designer should confidently and accurately move from usability to iteration are often vague or omitted entirely [1, 5, 12]. Perhaps, then, this hidden part of the UCD process should be more closely examined by design educators by taking a critical look at what professional designers do to make the initial move from usability to iteration as well as what they do to keep user-centered research findings at the forefront of the their concerns.

In the interim, however, we must try to help students avoid doing UCD that is unintentionally devoid of the findings from research. One such way is to make explicit that when decisions are advocated with the support of hypothetical outcomes (which are often simply the opinion of the designer) rather than with well-researched findings, the designers are no longer practicing UCD, but some sort of designer-centered design, where the desires of the designer, rather than the user, lead the development of a product. One such way to ensure that students are practicing UCD is to require documentation that explains why each design decision was made in the way it was and what from the usability testing supports that decision. By forcing novice designer to make explicit the roots of their decisions in educational settings, they will perhaps be able to more readily identify their own decision-making biases. Additionally, we must also help novice designers determine to what extent designer intuition can be incorporated into the design process without necessarily negating the UCD process. Indeed, recent research has pointed out that designer intuition has been undervalued as compared to more definitive research models [6, 17].

Conclusions

This case study is limited first and foremost by the fact that it studies only one group of novice designers. Second, I observed only the spoken exchanges from the members of the group, and thus did not account for written exchanges (such as email or memos) that could have contained more references to usability study. Third, and importantly, I make the assumption that the usability findings must be contained within the group's language in order for it to be present in the group's consciousness. However, conflicting information exists on this topic. Some researchers believe that there is a relationship between the language a person speaks and the way the person understands the world and behaves in it [18]. Therefore, in this case, the absence of usability results in the designers' language would suggest the absence of usability results in the designers' "world." Conversely, advocates of tacit knowledge would say that the lack of stated information does not indicate that it does not exist [3].

However, even if tacit knowledge is given its due, the lack of usability in the form of fact-based evidence is startling, and would seem to suggest that this group is doing something akin to but not quite UCD. Clearly, more information on both the activities of professional and novice designers, as well as role of usability instruction in the education of novice designers is needed in the future.

References

[1] Barnum, C. M. *Usability testing and research*. Longman, New York, 2002.

[2] Canary, D. J., Brossmann, B. G., & Seibold, D. R. Argument structures in decision-making groups. *The Southern Speech Communication Journal, 53* (1987), 18-37.

[3] Desouza, K. C. Facilitating tacit knowledge exchange. *Communications of the ACM, 46*, 6 (2003), 85-88.

[4] Dong, A. The latent semantic approach to studying design team communication. *Design Studies, 26* (2005), 445-461.

[5] Dumas, J. S., and Redish, J. *A practical guide to usability testing*. Intellect Books, Portland, Ore, 1999

[6] Fallman, D. Design-oriented human-computer interaction. In *Proc. CHI 2003,* ACM Press (2003), 235-232.

[7] Johnstone, B. *Discourse analysis*. Blackwell, Malden, Mass., 2005.

[8] Karat, J. Evolving the scope of user-centered design. *Communications of the ACM, 40*(7), (1997), 33-38.

[9] Karat, J., Atwood, M. E., Dray, S. M., Rantzar, M., & Wixon, D. R. (1996). *User centered design: Quality or quackery?* In *Proc. CHI 1996,* ACM Press (1996), 161-162.

[10] Kuhn, D. *The skills of argument*. Cambridge UP, New York, 1991.

[11] Macdonald, N. Beyond human-centered design? *interactions, 12*, 2 (2005) 75-79.

[12] Mao, J.-Y. V., Karel; Smith, Paul W.; Carey, Tom. The state of user-centered design practice. *Communications of the ACM, 48*, 3 (2005), 105-109.

[13] Meyers, R. A., Seibold, D. R., & Brashers, D. Argument in initial group decision-making discussions: Refinement of a coding scheme and a descriptive quantitative analysis. *Western Journal of Speech Communication, 55,* (1991), 47-68.

[14] Rubin, J. *Handbook of usability testing: How to plan, design, and conduct effective tests*. Wiley, New York, 1994.

[15] Sugar, W. A. What is so good about user-centered design? Documenting the effect of usability sessions on novice software designers. *Journal of Research on Computing in Education, 33*, 3, (2001), 235-250.

[16] Toulmin, S. *The uses of argument*. Cambridge UP, New York, 1958.

[17] van Eemeren, F. H., Grootendorst, R., & Henkemans, F. S. *Fundamentals of argumentation theory: A handbook of historical backgrounds and contemporary developments*. Lawrence Erlbaum Associates, Mahwah, NJ, 1996.

[18] Whorf, B. L. *Language, thought, and reality; selected writings*. Technology Press, Cambridge, MA, 1956.

[19] Wolf, T. V., Rode, J. A., Sussman, J., & Kellogg, W. A. *Dispelling "design" as the black art of CHI*. In *Proc. CHI 2006,* ACM Press (2006), 511-530.