

Benchmarking the effectiveness of universal design

Gary Scott Danford, Michael Grimble, and Jordana Maisel

State University of New York at Buffalo, Buffalo, New York

ABSTRACT: Evaluating the effectiveness of universal design is typically done through case studies conducted by experts in field settings. The logistics and costs of such case studies, however, inevitably constrain not only user participation but also the environments studied - further compromising case studies' already limited generalizability and frustrating evidence-based universal design practice. This paper proposes a method for benchmarking the effectiveness of universal design that removes these constraints and enhances case studies' generalizability by moving them from the field to the internet.

Conference theme: New methodologies in architectural research

Keywords: universal design, effectiveness, benchmarking, evidence-based practice

1. EVIDENCE OF EFFECTIVENESS

The most direct evidence of the effectiveness of any environmental design is the degree to which the environment enables the performance of salient activities by its users. For an environment that purports to comply with the *Principles of Universal Design* (Connell, et al, 1997) that would mean enabling comparable activity performance by all users (Danford, 2001; 2003) – i.e., facilitating full participation, inclusion, integration and equality regardless of any limitations associated with users' age, size, gender, abilities or circumstance.

Consequently, one might presume that evidence of universal design's effectiveness for all users could readily come from case studies of those users' experiences with such environments. But case studies on universal design are often designed to serve other purposes – e.g., as demonstration projects to provide examples of universal design's claimed benefits or as hypothesis tests to evaluate and refine universal design theory. As valuable as those purposes are, however, the data from such case studies have limited utility for the development of the translational research required for *evidence-based practice* (Hamilton, 2006).

Even those case studies that are specifically designed to generate evidence that facilitates translation from research to design – e.g., *action research* to solve an existing situation-specific problem (Lewin, 1976; Susman, 1983) and *post-occupancy evaluations* (POEs) to provide “feed forward” to the next design iteration (Preiser, et al, 1988) – are limited because the type and quality of data such studies gather are not standardized. Because both the type and quality of data vary so from one study to the next, the development of a cumulative body of evidence through meta-analysis (Cornell & Mulrow, 1999) is frustrated.

With the development of a cumulative body of evidence

so impacted, there is little empirical basis for predicting an environment's effectiveness or even for comparing one environment's effectiveness against another's. Such data make it virtually impossible to benchmark the effectiveness of universal designs that evidence-based practice necessarily requires.

An equally large obstacle to such case studies' translation from research to evidence-based practice is their external validity. Case studies' constraints raise formidable obstacles to the generalizability of the evidence they gather (Groat, 2002). Even with action research and POEs, pragmatic issues of logistics and costs inevitably work to restrict the users involved to small numbers and/or to delimit the environments that are available for study.

And, of course, case studies of universal design innovations that do not yet exist present their own unique obstacles to translation from research to evidence-based practice. Even when issues about internal validity that inevitably arise when such proposed environments are simulated (Catalano & Arenstein, 1993) can be satisfactorily addressed, threats to external validity attributable to yet additional issues of logistics and costs associated with developing and using simulations ultimately remain.

The most fundamental challenge to the generalizability of case study data, however, is that every case studied is inevitably confounded by unique situation-specific factors. Consequently, simply applying evidence gathered by single case studies to other user-environment situations is inherently problematic. At best, such applications run the risk of unintended consequences; at worst, they yield poor *user-environment fit* (Gifford, 2007) and/or maladaptive outcomes.

Evidence-based universal design practice requires the development of a cumulative body of evidence on environments' effectiveness in providing comparable

user-environment fit for all users. Research data on such effectiveness mined from diverse users' experiences with varied environments are required for an evidence-based practice of universal design that enables full participation, inclusion, integration and equality for everyone.

2. PROPOSING A NEW STANDARD

While traditional case studies on universal design pose numerous obstacles to conducting the translational research that evidence-based practice requires, it is possible to hurdle those obstacles by developing a new standard.

This new standard would have to produce case studies on universal design that (1) do not restrict user participation or delimit the environments that can be studied, (2) gather uniform data on diverse users' experiences with all environments, (3) generate evidence of environments' actual and relative effectiveness, and (4) strengthen the generalizability of that evidence to other user-environment situations.

To hurdle the obstacle posed by case studies' logistic- and cost-based constraints on user participation and the environments that can be studied, the traditional practice of bringing users to (or relying on users already present in) the environments to be studied could be reversed – e.g., bring the environments in simulated form to users via the internet. Not only would this enable almost unlimited user participation but also remove virtually all logistic- and cost-based limitations on the environments that can be studied. Expenses would be negligible compared to the prohibitive costs of the traditional case studies that would be required to generate comparable evidence.

To hurdle the obstacle posed by case studies' lack of uniformity in the type and quality of the data they gather, all case studies on universal design could gather the same evidence of user-environment fit for all users and environments – i.e., the incidence of problematic activities as a direct indicator of environments' effectiveness.

This standardization in the type and quality of data gathered also hurdles the obstacle posed by case studies' inability to generate the empirical evidence (as opposed to expert opinion) of effectiveness required by evidence-based universal design practice. By benchmarking the incidence of diverse users' problematic activities for each universally designed environment, empirical evidence of such environments' effectiveness can be gathered.

To hurdle the most fundamental obstacle faced by case studies – i.e., that every case studied is inevitably confounded by unique situation-specific factors – validated visual simulation methods such as line drawings (Stamps, 1993) could be used to bring all environments to all users. Supplemented by narrative descriptions to accommodate users who have vision conditions, the visual simulations would be able to control confounding factors and thereby enhance the external validity of case studies' evidence.

3. APPLYING THE PROPOSED STANDARD

An internet-based study of universal designs' effectiveness as indicated by activity performance begins with an accessible, W3C-complaint webpage that invites individual participants to complete anonymous online surveys about problems they have performing routine activities commonly encountered in everyday environments (e.g., public buildings, public streets, residential environments, etc.). While accessibility studies are typically focused on individuals who have mobility, sight, hearing or cognitive conditions, universal design studies also focus on individuals who do not have those conditions since the goal is to gather evidence of the designs' capacity for enabling full participation, inclusion, integration and equality for *everyone*.

The research design involves three phases. The first phase's *problematic activities surveys* establish the baseline incidence of problematic activities in each environment for each user group. The second phase develops *universal design proposals* for each environment intended to remedy those problematic activities. And the third phase's *design effectiveness surveys* then evaluate those design proposals' effectiveness by benchmarking their reduced incidence of problematic activities for each user group. All surveys are presented in accessible formats to enable ready use even to novice users of screen readers as well as individuals who have difficulty using a mouse or touch pad.

All participants begin by giving *informed consent* – i.e., agreeing to participate in the study as described on the website. An *Institutional Review Board for the Protection of Human Subjects* will typically consider this type of research *exempt* as long as it asks only innocuous questions that pose no risks to the anonymous participants. Even so, the general purpose of the study as well as the nature of participants' involvement completing the surveys are nevertheless explained – all without reference to *universal design* per se to avoid any biases, pro or con.

After providing informed consent, each participant's internet browser is automatically routed to a short *tell us about yourself questionnaire* where they provide standard demographic information about themselves. More importantly, they also provide information about (1) any sensory or functional conditions they have (e.g., limitations in mobility, vision, hearing, cognition, etc.) and (2) *how often* each condition affects their performance of routine activities by choosing the response option closest to their answer (e.g., *always* = 100% of the time, *usually* = 75% of the time, *sometimes* = 50% of the time, *rarely* = 25% of the time, and *never* = 0% of the time). These data are later linked to participants' responses to the subsequent *problematic activities surveys* and *design effectiveness surveys* to facilitate analyses specific to each user group.

The initial *problematic activities surveys* all ask participants *how often* they typically have a problem performing each of the listed activities by choosing the aforementioned response option closest to their

answer. Participants are then invited to explain their answers in terms of *why*.

The participants' *how often* and *why* responses are then analyzed to develop *universal design proposals* that reduce the incidence of the most problematic activities in each environment for all users.

These design proposals are then delivered back to participants through *design effectiveness surveys* that use both narrative descriptions and line drawings to communicate each proposal. All participants again begin by providing *informed consent* as well as the aforementioned *tell us about yourself* information in the event that their conditions and/or circumstances might have changed since completing the initial *problematic activities surveys*.

The *design effectiveness surveys* all begin by asking participants to reconfirm *how often* they typically have a problem performing each targeted activity. Participants are again invited to explain their answers in terms of *why*.

The *design effectiveness surveys* then present a neutrally-worded narrative description and simple line drawings of a proposed design that participants are told *might* improve their experience performing the activity. Line drawings are used to present the design proposals because they provide demonstrably valid simulations for the types of responses requested (Stamps, 1993) and can readily control potentially confounding situation-specific variables.

To test whether participants are actually responding to the narrative descriptions and line drawings, several design proposals meant to negatively impact performance of less problematic activities are purposely included in each *design effectiveness survey*.

Participants are then asked *how often* they would have a problem performing the activity in question if they were to encounter the proposed design by choosing the aforementioned response option closest to their answer. Participants are once more invited to explain their answers in terms of *why*.

4. PROBLEMATIC ACTIVITIES INDEX

To provide easily communicated and interpreted evidence of each universal design's effectiveness, a single index number is generated. This number, called the Problematic Activities Index (PAI) score, is based on (1) *how often* the participants' condition typically affects performance of routine activities in an environment and (2) *how often* the specific activity in question is problematic.

The PAI score is an index number that indicates how problematic an activity associated with an environment is on a scale from 0 to 100. The higher the score, the more problematic the activity and the less effective the environment.

The significance of a PAI score is always relative. The meaning of a score for an activity in one environment for one condition depends on how it compares to the scores for other activities, environments and/or conditions. For example, a low PAI score of 20 for an activity in one environment could nevertheless be worse than the PAI score for the same activity in a second environment, indicating that the second environment is more effective.

Activities' PAI scores enable the environments with which they are associated to be ranked on their effectiveness both within and across users' conditions. By knowing which activities present the greatest problems for various user groups in an environment, one can develop universal design solutions for that environment that improve participation, inclusion, integration and equality for everyone.

A two step process generates the PAI score for an activity based on the percentage of users reporting that a condition *always, usually, sometimes* or *rarely* affects their performance of routine activities who also say that their performance of the activity in question is *always, usually, sometimes, rarely* or *never* a problem (see Figure 1).

STEP 1: The number of users saying that the activity in question is *always, usually, sometimes, rarely, or never*

	A	B	C	D	E	F	G	H	I	J	
1	STEP 1:		CONDITION		CONDITION		CONDITION		CONDITION		1
2			Always	Always %	Usually	Usually %	Sometimes	Sometimes %	Rarely	Rarely %	2
3	ACTIVITY	Always	#	C3/A8*100	#	E3/A8*100	#	G3/A8*100	#	I3/A8*100	3
4		Usually	#	C4/A8*100	#	E4/A8*100	#	G4/A8*100	#	I4/A8*100	4
5		Sometimes	#	C5/A8*100	#	E5/A8*100	#	G5/A8*100	#	I5/A8*100	5
6		Rarely	#	C6/A8*100	#	E6/A8*100	#	G6/A8*100	#	I6/A8*100	6
7		Never	#	C7/A8*100	#	E7/A8*100	#	G7/A8*100	#	I7/A8*100	7
8	SUM(C8+E8+G8+I8)		SUM(C3+C4+C5+C6+C7)		SUM(E3+E4+E5+E6+E7)		SUM(G3+G4+G5+G6+G7)		SUM(I3+I4+I5+I6+I7)		8
9	STEP 2:		ACTIVITY		CONDITION		PROBLEMATIC ACTIVITIES INDEX SCORE				9
10			How Often	Multiplier	How Often	Multiplier					10
11			Always	100%	Always	100%	SUM(D3*D11*F11)+(F3*D11*F12)+(H3*D11*F13)+(J3*D11*F14)				11
12			Usually	75%	Usually	75%	SUM(D4*D12*F11)+(F4*D12*F12)+(H4*D12*F13)+(J4*D12*F14)				12
13			Sometimes	50%	Sometimes	50%	SUM(D5*D13*F11)+(F5*D13*F12)+(H5*D13*F13)+(J5*D13*F14)				13
14			Rarely	25%	Rarely	25%	SUM(D6*D14*F11)+(F6*D14*F12)+(H6*D14*F13)+(J6*D14*F14)				14
15			Never	0%			SUM(D7*D15*F11)+(F7*D15*F12)+(H7*D15*F13)+(J7*D15*F14)				15
16							SUM(G11+G12+G13+G14+G15)				16
	A	B	C	D	E	F	G	H	I	J	

Figure 1: Problematic activities index formulas

STEP 1:		CONDITION		CONDITION		CONDITION		CONDITION	
		Always	Always %	Usually	Usually %	Sometimes	Sometimes %	Rarely	Rarely %
ACTIVITY	Always	125	44	0	0	22	8	0	0
	Usually	0	0	0	0	0	0	0	0
	Sometimes	22	8	0	0	65	23	0	0
	Rarely	0	0	0	0	0	0	0	0
	Never	23	8	0	0	27	10	0	0
284		170		0		114		0	
STEP 2:		ACTIVITY		CONDITION		PROBLEMATIC ACTIVITIES INDEX SCORE			
		How Often	Multiplier	How Often	Multiplier				
		Always	100%	Always	100%	48			
		Usually	75%	Usually	75%	0			
		Sometimes	50%	Sometimes	50%	10			
		Rarely	25%	Rarely	25%	0			
		Never	0%			0			
						57			

Figure 2: Problematic activities index for using stairs for 284 participants with mobility conditions affecting legs and feet

problematic are placed in columns C, E, G or I based upon the participants' reports of *how often* their condition typically affects performance of routine activities. These frequency counts are automatically converted into percentages to normalize the counts on a scale from 0 to 100 in columns D, F, H and J.

STEP 2: The percentages in STEP 1's columns D, F, H and J are then weighted twice by (1) first multiplying the percentage by *how often* performing the activity in question is a problem (i.e., *always* = 100%; *usually* = 75%; *sometimes* = 50%; *rarely* = 25%; *never* = 0%) and (2) then multiplying by *how often* their condition typically affects performance of routine activities (i.e., *always* = 100%; *usually* = 75%; *sometimes* = 50%; *rarely* = 25%). The formulas for weighting the percentages are listed in G11-15. The sum in G16 is the Problematic Activity Index score for the activity in question that indicates the effectiveness of the environment with which the activity is associated – i.e., the lower the activity's score, the higher the environment's effectiveness; the higher the activity's score, the lower the environment's effectiveness.

5. THE PAI IN APPLICATION

The method for benchmarking the effectiveness of universal design proposed in this paper is currently being applied in a federally funded research project that is studying the effectiveness of universal design in various environments.

Figure 2 contains preliminary data from that project that demonstrate how an activity's PAI score is generated. Two hundred eighty-four participants with mobility conditions affecting their legs/feet reported *how often* their typical experiences using stairs in public buildings were problematic. These participants' answers were limited to three response options (i.e., *always* = 100% of the time, *sometimes* = 50% of the time, and *never* = 0% of the time).

Among the 170 participants who reported that their

mobility condition *always* affects their performance of routine activities, 125 said that their typical experience using stairs in public buildings was *always* problematic, 22 said it was *sometimes* a problem and 23 said it was *never* a problem.

Among the 114 participants who reported that their mobility condition *sometimes* affects their performance of routine activities, 22 said that their typical experience using stairs in public buildings was *always* problematic, 65 said it was *sometimes* a problem and 27 said it was *never* a problem.

The resultant PAI score for this user group's typical experience using stairs in public buildings is 57 on a scale from 0 to 100. This establishes the baseline PAI score for this activity that a universal design proposal for stairs in public buildings would lower by reducing the incidence of this activity being reported as problematic by this user group.

Such PAI scores have numerous applications in evidence-based universal design practice. For example, PAI scores for activities within or across user groups can be readily computed to score universal design alternatives' relative effectiveness enabling comparable activity performance for everyone.

CONCLUSION

Logically, evidence of universal design's effectiveness should come from case studies of users' experiences with various environments. But traditional methods of conducting case studies on universal design are inconsistent with the requirements of evidence-based practice. These inconsistencies present obstacles to the development of the translational research that evidence-based practice requires. By adopting a new standard for the conduct of case studies on universal design, however, those obstacles can be hurdled.

This new standard creates internet-based virtual case studies on universal design that (1) do not limit user

participation or constrain the environments that can be studied, (2) gather uniform data on all users' experiences with all environments, (3) generate evidence of environments' relative effectiveness, and (4) strengthen the generalizability of case studies' evidence to other person-environment transactions. By adopting this standard, universal design can hurdle the traditional case studies' obstacles that otherwise will continue to thwart direct translation from research to evidence-based practice.

ACKNOWLEDGEMENT

The method for benchmarking the effectiveness of universal designs described in this paper was developed in part with funding from the National Institute on Disability and Rehabilitation Research (NIDRR), U.S. Department of Education, through the Rehabilitation Engineering Research Center on Universal Design and the Built Environment, a partnership between the Center for Inclusive Design and Environmental Access (IDEA) and the Ontario Rehabilitation Technology Consortium (ORTC).

REFERENCES

Catalano, R. & Arenstein, W. 1993. Strategies of environmental simulation: theoretical, methodological and policy issues. In R. Marans & D. Stokols (Eds.) *Environmental Simulation: Research and Policy Issues*. New York: Plenum Press. 303-313.

Connell, B., Jones, M., Mace, R., Mueller, J., Mullick, A., Ostroff, E., Sanford, J., Steinfeld, E., Story, M., & Vanderheiden, G. 1997. *The Principles of Universal Design: Version 2.0*. Raleigh, NC: The Center for Universal Design.

Cornell, J. & Mulrow, C. 1999. Meta-analysis. In H. Adèr & G. Mellenbergh (Eds.) *Research Methodology in the Social, Behavioral and Life Sciences*. London: Sage. 285-323.

Danford, G. 2001. Principles of universal design. In G. Danford & B. Tauke (Eds.) *Universal Design New York*. New York: New York City Mayor's Office for People with Disabilities. 19-24.

Danford, G. 2003. Universal design: evaluation of a model building by persons with vision, hearing and mobility impairments. In A. Horowitz & C. Stuen (Eds.), Special Issue on Aging and the Senses, *Generations*. Vol. XXVII, No.1 (Spring), 91-94.

Gifford, R. 2007. Designing more fitting environments. *Environmental Psychology: Principles and Practice*. Coville, WA: Optimal Books, 523-558.

Groat, L. 2002. Case studies and combined strategies. In L. Groat & D. Wang (Eds.) *Architectural Research Methods*. New York: John Wiley & Sons. 341-373.

Hamilton, D. 2006. Four levels of evidence-based practice. *The AIA Journal of Architecture*. January. (http://www.aia.org/nwsltr_aiaj.cfm?pagename=aiaj_a_20041201_fourlevels).

Lewin, K. 1976. *Field Theory in Social Science*. Chicago: University of Chicago Press.

Preiser, W., Rabinowitz, H., & White, E. 1988. *Post-Occupancy Evaluation*. New York: Van Nostrand Reinhold.

Stamps, A. 1993. Simulation effects on environmental preferences. *Journal of Environmental Management*. 38, 115-132.

Susman, G. 1983. Action research. In C. Morgan (Ed.) *Beyond Method: Strategies for Social Research*. Newbury Park, CA: Sage Publications.

