

Methodological Issues in Research on Web-Based Behavioral Interventions

Brian G. Danaher, Ph.D. · John R. Seeley, Ph.D.

Published online: 6 October 2009
© The Society of Behavioral Medicine 2009

Abstract

Background Web-based behavioral intervention research is rapidly growing.

Purpose We review methodological issues shared across Web-based intervention research to help inform future research in this area.

Methods We examine measures and their interpretation using exemplar studies and our research.

Results We report on research designs used to evaluate Web-based interventions and recommend newer, blended designs. We review and critique methodological issues associated with recruitment, engagement, and social validity.

Conclusions We suggest that there is value to viewing this burgeoning realm of research from the broader context of behavior change research. We conclude that many studies use blended research designs, that innovative mantling designs such as the Multiphase Optimization Strategy and Sequential Multiple Assignment Randomized Trial methods hold considerable promise and should be used more widely, and that Web-based controls should be used instead of usual care or no-treatment controls in public health research. We recommend topics for future research that address participant recruitment, engagement, and social validity.

Keywords Internet · Web-based · Interventions · Design · Measurement · Recruitment · Engagement

Introduction

The literature describing Web-based interventions is nascent but growing very rapidly. A review and critique of shared methodological issues may present a useful guide to help in our understanding of this research and in shaping future inquiry. For the current review, we focus on publications that describe the use of a Web-based intervention. We examine this research from the perspective of existing research models, some of which may need to be adapted to map onto issues pertinent to Web-based intervention research. For example, we consider the role of blended designs that accommodate the difficulty in differentiating between large-scale Web-based efficacy trials from carefully controlled Web-based effectiveness trials. And we explore issues surrounding the selection of optimal control/comparison conditions. The paper also examines methodological issues associated with participant recruitment (representativeness, recruitment messages, and modalities), participant engagement (participant exposure and skills practice), and social validity (consumer acceptance). While the scope of the paper does not enable us to provide an exhaustive list of citations, we nonetheless have cited exemplar publications and we have presented relevant portions of our own research.

Research Design

The design of Web-based behavioral intervention research can be considered a special case of traditional behavioral research designs and methods that have been developed and applied long before the advent of the Internet. In this section, we provide a brief summary of the design and evaluation methods that are particularly relevant to Internet

B. G. Danaher (✉) · J. R. Seeley
Oregon Research Institute,
1715 Franklin Boulevard,
Eugene, OR 97403, USA
e-mail: briand@ori.org

intervention research. We begin with a brief overview of the stages of research and the types of program evaluations that are relevant to these research stages, we highlight study designs of particular relevance to Web-based intervention research, and we discuss issues related to control/comparison groups.

Stages of Research and Program Evaluation

The Stage Model of Behavioral Therapies Research [1] articulates three progressive stages of development and evaluation of behavioral interventions. This model is especially relevant to Web-based intervention research given its goals of encouraging innovation and facilitating widespread use of empirically validated behavioral programs. In addition to the stages of research paradigm, the USDHHS Science Panel on Interactive Communications and Health [2] adapted a program evaluation model that provides guidance to systematically obtaining information to improve the design, implementation, engagement, adoption, redesign, and overall quality of an intervention. The type of evaluation depends on the stage of research and the research goal. Formative evaluation is typically used in stage I to assess the nature of the problem behavior and the needs of the target population in order to inform intervention design and program content. Process evaluation is used in that stage to monitor the administrative, organizational, and operational aspects of the intervention. Outcome evaluation is then used to determine whether the intervention can achieve its intended effects under both ideal conditions (stage II efficacy research) and more real-world conditions (stage III effectiveness research).

Stage I research includes activities related to the (a) treatment method, (b) therapist involvement, (c) participant selection, and (d) design and analysis. As a special case of behavioral therapies, Web-based intervention research shares many of the elements of the stage model. Stage I research is considered an iterative process of intervention development and evaluation [3] which serves as a vital link between feasibility research and subsequent research stages. Formative evaluation may include (a) needs assessment; (b) evaluation planning; (c) communication strategies; and (d) determination of specific system requirements, features, and user interface specification. Stage II research typically employs experimental designs (e.g., randomized controlled trials [RCT]) to evaluate the efficacy of behavioral interventions that have shown promise during the stage I research. Because internal validity is the major emphasis, stage II trials typically involve fairly homogeneous samples and settings.

This final stage of research focuses on the transportability of efficacious interventions to real-world contexts and large-scale dissemination. The key issues include generalizability,

implementation, cost-effectiveness, and social validity (acceptability to end users, program adopters, health care providers, policy makers). Evaluation activities and methods are generally the same as found in stage II efficacy research, but with more emphasis on participatory process evaluation given the inclusion of more heterogeneous subjects and settings. Research in this stage is guided by the RE-AIM framework [4] which delineates measures and reporting that gauge the generalizability, reach, and dissemination of intervention efforts.

Design Considerations

The stage model was designed to facilitate the sequential development of promising interventions from the inception of clinical innovation to dissemination. However, research does not always follow a serial path from efficacy to effectiveness and, in trying to impose such distinct stages, innovative Web-based interventions may be overlooked [5]. For example, experimental control may be less rigorous in stage II trials studying Web-based interventions that occur in real-world contexts rather than in more highly controlled research settings. As we describe in the following section, it is important to consider whether the research design is blended or hybrid, if it is adjunctive to other treatments or programs, and/or whether it uses mantling or dismantling evaluation approaches.

Blended Designs

In an effort to facilitate greater emphasis on issues such as utility, practicality, and cost earlier in the evaluation of promising therapies, Carroll and Rounsaville [6] propose a hybrid model to link efficacy and effectiveness research. The hybrid model retains essential features of efficacy research (randomization, use of control conditions, independent assessment of outcome, and monitoring of treatment delivery) while expanding the research questions to also address issues of importance in stage III effectiveness studies. Such dissemination issues include heterogeneous settings and participants, cost-effectiveness of the intervention, training issues, and patient and clinician satisfaction.

Other blended designs that address similar issues of external validity and dissemination include *practical clinical trials* or PCT [7] and *pragmatic randomized controlled trials* or P-RCT [8]. Glasgow [9] has advocated for the use of PCTs in eHealth intervention research that address key dissemination issues such as acceptability and robustness of programs across heterogeneous subgroups and settings. P-RCTs have similar aims of minimizing exclusion/inclusion criteria to reflect the heterogeneity of the targeted population. However, PCT and P-RCT designs differ with respect to the selection of comparison conditions. Glasgow [9] recommends

that Web-based intervention PCTs employ—as controls—reasonable treatment alternative intervention choices rather than no-treatment or usual care comparison groups. P-RCTs tend to deal with each treatment as a black box and usually are not concerned with understanding the active ingredients within the box [8].

Adjunctive Designs

Studies have tested the adjunctive benefits of combining Web-based interventions with other types of treatments. In some instances, the Web-delivered program is adjunctive to the other treatment whereas, in other instances, the other treatment modality is adjunctive to the Web-based. Examples of Web-based adjuncts to other treatments include online maintenance programs following completion of a clinic-based program [10, 11], personalized e-mail coaching as adjuncts to self-help print materials [12], and an online cognitive behavioral program to supplement nicotine replacement therapy [13]. Alternatively, a number of studies have examined the value of using Web-based intervention supplemented with personalized, non-automated e-mail counseling [12, 14], synchronous online group counseling [15, 16], and telephone counseling [12, 17]. Such adjuncts to Web-based programs have been employed to facilitate participant skill acquisition and practice, to provide additional support, and to enhance participant motivation and engagement. However, unless an analysis of the adjunctive components of the intervention is conducted, the active ingredients of the Web-based program cannot be discerned from the effects of the adjunctive components.

Dismantling and Mantling Designs

A growing number of researchers have used mantling/dismantling designs in order to isolate the effect of particular Web-based program components [18]. Mantling designs are used to determine the active components or features of a multicomponent intervention early in the development cycle (e.g., late stage I or early stage II), whereas dismantling designs are employed after an intervention package has been proven to be efficacious during stage II. For example, Wangberg [19] compared a Web-based diabetes self-care intervention with and without a component designed to encourage greater behavioral self-efficacy. Tate and colleagues [20] dismantled the effects of e-mail counseling by giving participants in the control group all other intervention ingredients. Strecher and colleagues [13] tested two Web-based programs, one of which provided tailored program content. Thompson and colleagues [21] evaluated the effects of immediate versus delayed financial incentives to encourage participant visits to a Web-based youth obesity

prevention program. Both Clarke and colleagues [22] and Lenert and colleagues [23] examined the effects of e-mail on outcome in a Web-based intervention.

A limitation to traditional approaches for evaluating the active components or features of an intervention described above is the number of intervention arms that can be compared within a randomized trial. To address this limitation, Collins and colleagues [24] have proposed two new methods for mantling and evaluating Web-based interventions that address this limitation: the Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART). The MOST method consists of three phases: (1) screening potentially active components of a black box intervention, (2) evaluating the effects and identifying the optimal dose of the most relevant components identified in the screening phase, and (3) confirming the optimized set of the components through an experimental evaluation (e.g., RCT, P-RCT, or PCT). In the first and second phases, MOST incorporates a fractional factorial experimental design which allows for testing theory-driven combinations of intervention components rather than crossing all possible combinations. The SMART method is another tool for evaluating and refining the potentially active treatment components during the second phase of MOST. The SMART method employs a randomized experimental design for building time-varying adaptive interventions and addresses questions regarding the sequencing of components, tailoring, and information architecture (e.g., menu of treatment options versus assignment of treatment components).

Strecher and colleagues [25] have recently employed the MOST method to identify the active components of a Web-based smoking cessation program plus nicotine patch. In what could be characterized as both a blended design and a pharmacologic-adjunctive design, these researchers attempted to recruit the entire population of smokers from two health maintenance organization (HMO) settings as the sampling frame for the study. Smokers ($n=1,848$) were randomized to one of 16 experimental arms in which each arm comprised a different combination of five psychosocial and communication Web-based components. Seven-day smoking abstinence at 6-month follow-up was related to high-depth tailored success stories and a highly personalized message source. This study illustrates that methodological innovations in Web-based intervention research have the potential to provide considerable cost savings for more rapid prototype testing and identification of the most active intervention components [26].

Comparison Group Considerations

As research on Web-based intervention grows, certain trends are emerging with regard to the type of comparison

or control conditions being used in experimental trials. In selecting an appropriate comparison or control condition, the researcher must weigh pragmatic, ethical, and experimental control considerations. With respect to pragmatic considerations, blended designs have been employed to maximize the external validity of the study and relevance to clinical practice and dissemination. This may involve randomization of heterogeneous subgroups to no-treatment or usual care comparison conditions in the case of pragmatic RCTs or to alternative effective treatment conditions for practical clinical trials. Ethical considerations regarding control selection include the level of risk of the experimental treatment versus the comparison condition and the availability of other effective treatments. Experimental control issues include the ability to (a) differentiate the efficacy of an intervention from the natural course of the targeted behavior, (b) determine whether the effects of a treatment exceed the effects of nonspecific or common factors, and (c) ascertain that the intervention effects are not caused primarily by participant expectancies regarding behavioral change. Hence, a critical consideration when selecting a control condition is what can be held constant because any differences between the experimental arms could represent alternative explanations for outcome differences that are obtained [27]. Another important consideration concerns the aim of the intervention and issues related to the external validity of the study. That is, a Web-based intervention aimed at improving outcomes for primary care patients might employ a clinic-based usual care comparison condition whereas a Web-based comparison condition may be more appropriate for a more general public health prevention program.

Clinic or Standard/Usual Care Comparison Conditions

A number of researchers have examined the impact of Web-delivered interventions relative to standard or usual care such as clinical treatment [11, 28–32] or self-help manuals [33]. Standard or usual care designs address how a relatively newer Web-based treatment compares to a more traditional treatment approach when each intervention takes full advantage of its respective medium [30]. A persuasive case can be made for the merits of comparing new treatments to established treatments [18]. Major advantages for this type of design include addressing (a) ethical concerns about providing treatment to all research participants (assuming usual care is effective) and (b) practical issues regarding participant recruitment as well as informing on dissemination. But when standard care is delivered exclusively in a clinic or classroom, it is important to acknowledge that results from participants who presumably agree to be assigned to either treatment modality may not be representative of the much larger audience of potential users for whom a Web-based

intervention represents the only feasible or attractive option. Furthermore, if the standard or usual care has not been proven to be effective or involves weak dosage, such comparison groups may only control for the passage of time and participant expectancies, but they may not adequately address other confounds such as demand characteristics.

Web Comparison Conditions

Many studies have used credible Web-based control/comparison conditions. Such comparison conditions have an advantage of potentially controlling for demand characteristics and participant expectancies better than standard or usual clinical care. In addition, a Web control condition can have practical benefits with respect to ease of implementation and broad recruitment base, hence, potentially reaching more diverse subgroups and providing greater generalizability of study findings. One example is the use of a *basic* information website control condition that presents facts about the target behavior and possibly some treatment recommendations [13, 14, 34–36]. The *enhanced* condition typically contains tailored program content, interactive features that help the participant clarify motivation and/or make and then follow-through with a behavioral plan, video models (testimonials), and messages about behavioral practice. Basic conditions could be considered *usual care* controls because they contain content that individuals could typically find via searching the Web, but they are not expected to be as effective as an *enhanced* intervention [37, 38]. The extent and type of content offered in *basic* website controls vary considerably across studies as researchers have balanced the risk of providing sufficient content to be credible while also ensuring that the enhanced condition will show a treatment effect via specific active mechanisms. If an external website is used as the control condition, then program usage data may be limited to tracking logins via a portal rather than measuring usage duration, specific webpage views, or website components used.

No-Treatment and Waiting List Controls

Some studies have used a waiting list/no-treatment control [12, 38–40] which offers pragmatic advantages to the researcher but has only limited value in being able to provide explanatory information about the processes (e.g., effects of nonspecific factors) and the magnitude of change [41, 42]. In clinical settings, the waiting list may be better described as usual care comparison but in standalone fully automated Web interventions, the waiting list translates into a no-treatment condition. While cross-over designs are sometimes used (control participants eventually offered the

active treatment), their results are typically not reported. Because no-treatment comparison groups only control for the passage of time (i.e., the natural course of the behavior), their utility may be limited to the more formative stages of the intervention development and evaluation process.

Recruitment

As is true in any intervention research, it is important to avoid scenarios in which too few participants are enrolled [43, 44]. Some published reports of Web-based interventions have cited highly successful recruitment efforts [45], and others have described how recruitment plans had to be adjusted in reaction to unsatisfactory initial results [5, 46]. In this section, we examine some of the key themes associated with recruitment including representativeness, recruitment messages, and recruitment modalities.

Denominator Considerations

When a Web-based intervention uses an open recruitment approach, any effort to compare enrolled participants to the eligible population (i.e., to define the denominator) is fraught with conceptual and practical problems. When recruitment is limited to members of a known population (e.g., members of the military or a HMO [22, 47]), then it becomes more feasible to compare the characteristics of recruits to nonparticipants, as recommended in the RE-AIM framework [9]. One possible concern focuses on the representativeness of study participants (i.e., external validity of results). Using the broadest view, this concern would focus on differences in Internet access associated with the so-called *digital divide* [48] which Nielsen [49] cogently considers as having three parts: economics, usability (to accommodate lower computer literacy), and empowerment (not using Web tools to their full potential). Participants in most Web-based research are probably elevated on the continuum of low–high computer literacy and computer self-efficacy. Greater variability on these dimensions could be expected in studies that purposefully target their recruitment to underserved computer/Internet users and provide the novice users with computers and Internet access [50]. This research is best considered a special case because the evaluation of these Web-based interventions is confounded by learning how to use a computer, the Internet, and learning to navigate program content in a Web browser.

Marketing Message Attributes

Researchers can vary the messages they use to market their Web-based interventions in order to recruit a sample of

study participants having a sufficient size and characteristics consistent with study aims and the research design. For example, social marketing themes could be emphasized in order to appeal to an audience that shares psychographic characteristics [51]. Online mental health interventions could highlight their anonymity and the resulting lack of stigma [52]. Marketing messages might address prospective participants' sense of comfort in using new technology tools. The Technology Acceptance Model [53] posits that people are more likely to adopt and use new technology tools—like Web-based programs [54]—when they perceive the tools as being both useful and easy to use. Marketing messages that highlight program benefits could help distinguish between a Web-based intervention and the more familiar information-based websites [55]. The messages could help recruits make a more informed decision about their responsibilities as research participants and result in improving subsequent engagement and/or program completion. Accentuating the credibility of the source of the Web-based program can also be helpful. However, credibility is in the *eye of the beholder*. Results from one study [56] suggests that, compared to experts, online consumers are more influenced by information architecture and esthetic design when judging the credibility of health websites (including websites for the National Institutes of Health and Mayo Clinic).

The marketing messages for RCTs need to accurately reflect the program(s) being offered because it is safe to assume that prospective participants choose programs based on their preferences. Consider the scenario of a prospective participant who is attracted to a program because of a marketing message that describes the many benefits of a highly tailored, interactive online program. If that individual is subsequently assigned to a control condition that instructs study participants to use only print materials, then as a study participant that individual might be less engaged, complete fewer assessments, or even drop out because of the perceived inconsistency between *the offer* and *the result* in the marketing message.

Recruitment Modalities

The modality used to deliver the marketing message can enhance *website exposure* and *website visibility* which results in greater recruitment. Some researchers [45, 46, 57] have successfully used a combination of paid advertising and unpaid (earned media) announcements via online resources (e-mail, search engines, affiliate websites, online communities) and more traditional media (newspapers, newsletters, flyers, radio, and TV), direct mail, and phone. Because the intervention is Web-based, it is reasonable to assume that recruitment might be enhanced through advertising on online search engines (e.g., using Google AdWords). During

recruitment for our recent RCT of a Web-based smoking cessation intervention [46], we paid for our ads to have high visibility in Internet search engine results in order to successfully recruit more than 2,000 participants from across the US (70% of participants came from Google ads and 20% from Yahoo ads). Figure 1 shows the functional relation between these ads and participant recruitment using an ABAB reversal design (ads displayed during weeks 1–8 and week 10).

Another modality that has received little mention in published reports involves the target of the marketing effort: where respondents are directed to get more information. Often, the target involves a project marketing website that describes features of the research, highlights the credibility of the content and source, answers common questions (FAQs), and provides a method for screening and enrollment. A variation on this theme is the *guest website* (or *guest* portion of the project marketing website) targeted to stakeholders who can facilitate participant recruitment (e.g., spouses and significant others, teachers, state health officials). The effectiveness of project marketing websites is likely to be influenced by how well they describe the features of the available programs as well as their functionality (e.g., use of text, multimedia, interactivity) and their esthetic design. Marketing campaigns could also encourage prospective participants to call or write to a research project office to obtain more information.

A number of studies have examined recruitment modality. For example, Thompson and colleagues [21] reported that participants recruited via broadcast media visited a Web-based intervention more often than did participants recruited via more traditional methods. Severson and colleagues [36] reported that recruitment modality was unrelated to outcome. Glasgow and colleagues [47] found that more members of a diabetes registry enrolled in a Web-based weight loss intervention after they received personalized letters from a senior health plan physician that accompanied the publication of articles in HMO member newsletters. In a within-

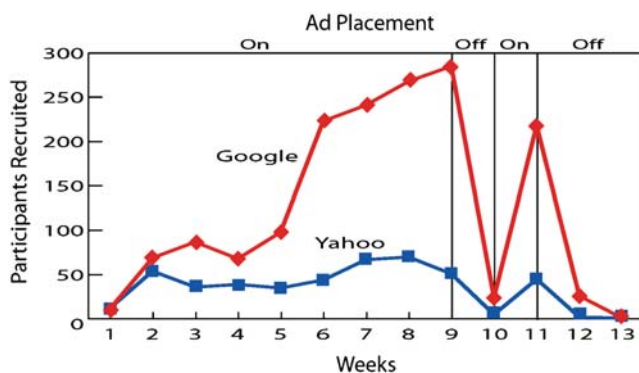


Fig. 1 Positive effect of online search engine ad placement on recruitment of participants to an online smoking cessation program [46]

modality test, Alexander and colleagues [58] found that different monetary enrollment incentives in mailed invitations increased recruitment to a Web-based nutrition intervention trial. Finally, some studies have described the cost per recruit based on the fact that different recruitment modalities vary in terms of the costs for materials, staff time, and media expenses [45].

Engagement

Within the constraints of the website's information architecture [59], each participant is able to decide how much of the program and how long and often they want to access it. They can also decide how much they actually practice what they learn. There is accumulating evidence that many participants of Web-based interventions exhibit less program engagement than program developers envisioned [60]. Research on Web-based interventions typically includes measures of engagement. There are multiple benefits to measuring engagement, including ensuring program usability, determining what participants use in the context of what they are offered [61], and identifying active ingredients that help to explain any observed treatment effect.

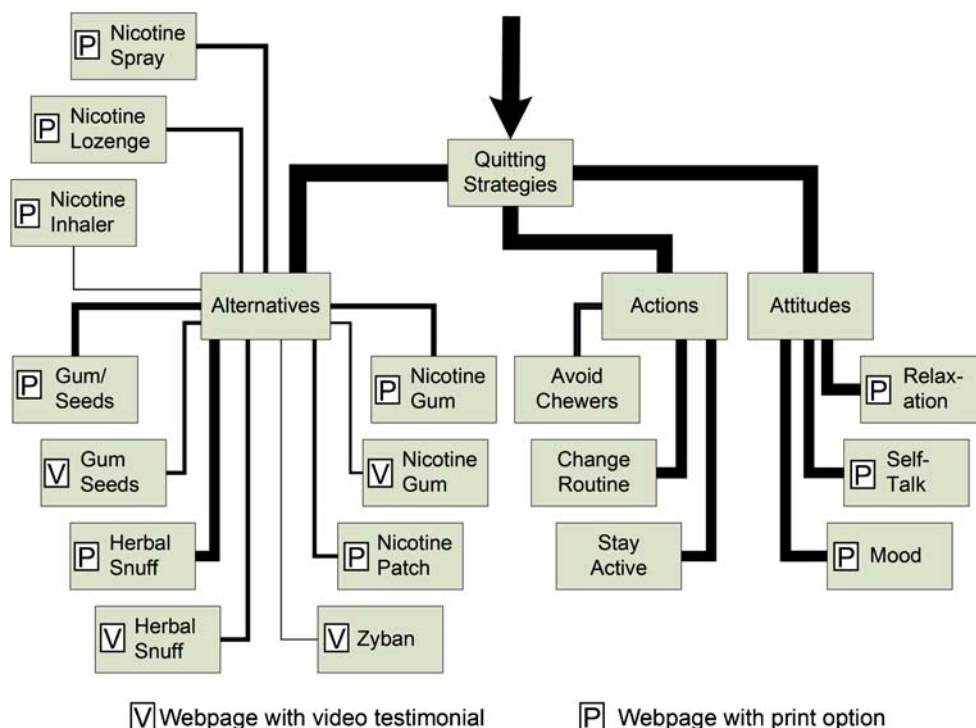
Unobtrusive Measures of Participant Exposure

Depending upon the particular Web server technology, it is possible to unobtrusively and objectively measure participant use of a Web-based program by analyzing server log files that are typically created for each website transaction [62]. Another approach involves using scripting software and webpage tagging to obtain more detailed usage data recorded on a server-side database [63]. Data on participant exposure to a Web-based intervention has become an expected ingredient in published reports [63].

Webpage Viewing

Measures of unique webpage viewing attempt to address the reasonable concern about whether a participant has seen *enough* of the program content. It is important to acknowledge that webpage viewing measures can be difficult to interpret since they can be greatly influenced by the information architecture of a Web-based intervention which may require users to view portions of the program in a predefined order [59]. For example, Fig. 2 shows a traversal paths diagram of webpage viewing by participants in the Web-based ChewFree smokeless tobacco cessation program [36]. In this example, webpages on *Quitting* are depicted as boxes and the proportion of participants who viewed the webpage at least once is

Fig. 2 Traversal paths of webpage viewing by participants in the ChewFree RCT [36, 63] (line thickness proportionate to number of participants who viewed each webpage)



depicted by the thickness of the line connecting the boxes (thicker lines indicate a greater proportion of participants accessed the next webpage in the flow). As this diagram shows, the top-down hierarchical information design caused participants to view webpages higher in the structure in order to view webpages lower in that structure. Webpage viewing can, therefore, be an artifact of the information architecture of the program. Graphical portrayals of paths of use through a website have considerable value for formative testing and usability testing, although it is not as clear how they can best be used to compare conditions or summarize results across studies.

The emergence of software tools to develop rich Web-based applications allows program content to change without refreshing the webpage. It is also increasingly common for webpages to include user-activated videos and animations. As a result, the value of webpage views as a Web metric has been diminished because it is no longer possible to assume that webpages display a fixed amount of information. Nielsen/NetRatings, a major online measurement service, recently announced that it ceased using webpage views as its index of site usage [64]. While many current Web-based interventions may not be considered rich applications or deliver content using videos, we nonetheless believe it is important to acknowledge the challenge to keep objective measures of engagement consistent with the inevitable and rapid changes in Web technology.

Viewing Duration

Program exposure can also be measured in terms of the amount of time users spend when accessing the Web-based program [65, 66]. Nielsen/NetRatings now ranks websites using total duration and average time of all visits. Measures of duration address the reasonable concern about whether a user has had enough time to be exposed to the content in the program. Time spent viewing a Web-based program can be affected by factors that are unrelated to measuring engagement [67]. For example, broadband users can more rapidly access program content than participants who use dial-up Internet service access [68]. Duration measures can be inflated for participants who engage in multitasking (e.g., accessing the Web-based intervention while concurrently viewing other websites, watching TV, or talking on the phone). In the extreme, the clock would continue to tick when a participant had left their computer completely unattended. While it is not feasible to control for multitasking, it is possible to try to minimize artifacts associated with unattended viewing by adopting an a priori decision rule in analyzing usage data such that any single webpage with at least 30 min of continuous viewing should signal the end of the session, dropping or truncating the duration of that webpage from any calculations and redefining the end of the session as the end of the last allowable webpage viewed [63, 69]. Participant privacy can also be enhanced and hosting burden reduced by designing Web-based interventions to

automatically end sessions (visits) after 30 min of inactivity.

Program Visits

Many studies have reported unobtrusively measured server data on the number of website visits [70]—sometimes referred to as *website hits*. While visits would seem to be discrete events, it is important to note that operational definitions for a visit should be considered. For example, comScore Media Metrix, an online measurement service, defines visits in terms of returns to a website with a break of at least 30 min [64, 71]. In addition, periods of inactivity lasting 30 min can also be used to operationally define the end of a visit. Note that some studies [63] have examined the number of days on which at least one visit occurred and/or days since project enrollment on which at least a single visit occurred.

Composite Measures

Because of the aforementioned complexities in measuring webpage views, duration, and visits, some researchers have combined these data into composite measures. For example, the online ChewFree smokeless tobacco cessation trial [36, 63] used a composite exposure variable defined as the mean of the *z*-score transformation of the overall number of visits and duration. The analysis of Cobb et al. [65] of the Web-based QuitNet smoking cessation program used the product of the number of logins and duration per login (in minutes).

Skills Practice

Most observers would agree that, in order to achieve meaningful behavior change, a participant would need to engage in new behaviors in his/her regular routine (skills practice). Thus, it comes as no surprise that Web-based programs typically recommend skills practice. The extent to which participants engage in skills practice can be measured using questions on follow-up assessments that ask about the timing, the type, and the extent of behavioral practice. Some Web-based interventions provide online mechanisms for participants to log, rate, and/or provide commentary on practice episodes outside of interacting with the online program. Server data for these online features (diaries, logs, ratings, and commentary) provide proxy measures for skills practice. For example, Farvolden and colleagues [72] described how a Web-based intervention for panic attacks queried participants about the “homework they were assigned.”

It is also possible for a coach, consultant, or therapist to review the homework, as described in a report of a Web-

based cognitive behavioral intervention for social phobia [73]. The use of free-standing tools to validate self-report is another possibility. For example, data from a wrist-worn pedometer can be used in the evaluation of a Web-based physical activity program [74]. Additional technologies can be used to track skills practice activities that occur outside of the online visit. Hand-held devices (PDAs and cell phones with text messaging and interactive voice response [IVR]) might be used for ecological momentary *dialogs* that combine assessment and tailored program messages to encourage skills practice in participants' normal routines. Brendryen and colleagues [75] described a successful fully automated Web-based smoking cessation intervention that used e-mail and hand-held technology tools (IVR, SMS text messaging). Pushing the envelope even further, it might be possible for participants to use their cell phones or digital cameras to film and then upload brief videos that confirm in vivo practice of behavioral skills (e.g., [76]). Interpreting multifaceted measures of skills practice and then combining those data with the aforementioned measures of engagement represents a conceptual and practical challenge.

Relation to Outcome

Participants' exposure status could be described by sample statistics (e.g., mean or median split, groupings by quartiles) or on the basis of thresholds derived from clinical settings [61]. Some studies have excluded [63] or analyzed results separately [77] for participants who never visited the Web-based program to access content (some of whom nonetheless completed online assessments). Strecher and colleagues [13] described results of a Web-based smoking cessation program using three nested partitions of participant engagement and assessment completion: all participants who were randomized (baseline observation carried forward or missing = smoking), participants who logged on at least once, and participants who completed the follow-up assessment.

Engagement data may be an important underlying mechanism that explains a portion of the variance in observed treatment effects. In a recent paper [37], we examined program exposure and change in self-efficacy as putative mediators of the effect of treatment on outcome in the ChewFree RCT. Exposure was defined as the mean of the *z*-score transformations of visits and duration. Both self-efficacy change and exposure were found to be simple mediators, but only self-efficacy remained a mediator when the variables were tested for multiple mediations [78].

This is not to say that the hypothesized cause and effect relation is automatically strengthened by a dose–response relationship between program engagement and improved outcomes. First, the dose–response relationship is complicated by attrition. Christensen and her colleagues have used

the terms *e-attainer* [52] and *one hit wonder* [79] to describe participants who benefit from brief exposure to Web-based interventions—some of whom are likely to be included among dropouts. Looking only at *completers*—or imputing baseline levels for dropouts—potentially confounds the relationship. Secondly, the dose–response relationship might actually be measuring another attribute—possibly participant motivation—which overlaps with or is obscured in analyses by program engagement. Finally, it is prudent to consider the research evidence showing that more is not always better: sometimes doing more can have an iatrogenic impact. For example, Glasgow and colleagues [47] reported that engagement decreased as requirements on the participant increased, prompting the authors to conclude that “...adding components to a basic Internet-based intervention program can create adherence challenges.” Similarly, Lenert [80] concluded that adding a mood management component was potentially harmful to an online smoking cessation intervention.

Social Validation and User Satisfaction

Many studies have described relatively global measures of participant satisfaction [35], program relevance [81], and/or whether participants would recommend the program to others [46]. We believe that there is more utility in asking participants to provide this type of global rating of satisfaction rather than asking for satisfaction ratings of specific program features by name. One problem lies in the fact that participants may not think of a program module using the same labels as the researcher. In our experience, many participants rate the value of a specific module or feature when, in fact, server data indicates that they have not viewed it. We recommend that future research should focus additional attention on assessing the user experience [82] by measuring a program's overall usability [83] as well as usability dimensions specifically relevant to Web-based interventions.

Conclusions

A number of key methodological issues in research on Web-based interventions were addressed in this review. We conclude that many studies in this area are best described as having blended research designs. That is, given the ease with which Web-based interventions can be scaled up to reach diverse participant groups, the path from stage II efficacy research to stage III effectiveness research does not typically follow a serial progression as is often the case for traditional behavioral intervention research. Although not widely used as yet, innovative mantling designs such as the

MOST and SMART methods hold considerable promise because (a) Web-based interventions can be designed to vary their delivery of content to participants and (b) such designs can help us improve our understanding of the optimal combination and timing of most active program ingredients. Our review of comparison conditions concludes by recommending the use of Web-based controls in public health research in order to control for demand characteristics and participant expectancies. Clinic-based usual care comparison conditions would be more applicable when the Web-based intervention aims to improve outcomes for primary care patients.

We also conclude that increasing research attention and methodological sophistication should consider participant recruitment, participant engagement, and social validity (consumer acceptance). More research should focus on how recruitment is best accomplished (marketing message and modality) as well as on its noteworthy impact on representativeness and external validity. Issues surrounding screening and informed consent should be further explored. For example, we recommend that researchers test the utility of providing prospective participants with a more detailed explanation of study participant roles and responsibilities, and we allude to the likely beneficial use of multimedia as a part of the informed consent procedure. But we advise caution regarding the use of stringent run-in requirements that could have the effect of culling out all but highly motivated study participants.

Researchers routinely recommend that future research needs to identify approaches to encourage greater participant engagement. Data are also needed to clarify what constitutes a sufficient extent of participant exposure—especially because participant commitment might vary by target behavior, and it might be a proxy for motivation. Research should also more fully explore the utility of using composite measures of engagement/exposure that incorporate multiple program usage metrics. Similarly, more analyses, including tests of mediation, are needed to clarify the relation between skills practice and resulting changes in target behavior(s). Finally, research needs to include a greater emphasis on social validity or the extent that participants like the program to which they were assigned. We believe that these topics—and their component themes—should be considered critically important opportunities as well as essential methodological features of research in this fledgling field.

Acknowledgments This work was supported, in part, by grants CA118575 and CA84225 from the National Cancer Institute. We would like to extend our thanks to the following individuals: H. Garth McKay, who made important contributions to an earlier version of this paper; Edward Lichtenstein, who reviewed multiple drafts and participated in brainstorming sessions; and both Deborah Toobert and Stephen Boyd, who reviewed earlier report drafts.

References

1. Onken LS, Blaine JD, Battjes RJ. Behavioral therapy research: A conceptualization of a process. In: Hennigler SW, Amentos R, eds. *Innovative Approaches From Difficult-To-Treat Populations*. Washington: American Psychiatric Press; 1997: 477–485.
2. Eng TR, Gustafson DH, Henderson J, Jimison H, Patrick K. Introduction to evaluation of interactive health communication applications. Science Panel on Interactive Communication and Health. *Am J Prev Med*. 1999; 16(1): 10–15.
3. Rounsaville BJ, Carroll KM, Onken LS. A stage model of behavioral therapies research: Getting started and moving on from Stage I. *Clin Psychol Sci Pract*. 2001; 8(133): 142.
4. Glasgow RE, Vogt TM, Boles SM. Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *Am J Public Health*. 1999; 89(9): 1322–1327.
5. Dansky KH, Thompson D, Sanner T. A framework for evaluating eHealth research. *Eval Program Plann*. 2006; 29: 397–404.
6. Carroll KM, Rounsaville BJ. Bridging the gap: A hybrid model to link efficacy and effectiveness research in substance abuse treatment. *Psychiatr Serv*. 2003; 54(3): 333–339.
7. Tunis SR, Stryer DB, Clancy CM. Practical clinical trials: Increasing the value of clinical research for decision making in clinical and health policy. *JAMA*. 2003; 290(12): 1624–1632.
8. Hotopf M, Churchill R, Lewis G. Pragmatic randomised controlled trials in psychiatry. *Br J Psychiatry*. 1999; 175: 217–223.
9. Glasgow RE. eHealth evaluation and dissemination research. *Am J Prev Med*. 2007; 32(5 Suppl): S119–S126.
10. Harvey-Berino J, Pintauro SJ, Gold EC. The feasibility of using Internet support for the maintenance of weight loss. *Behav Modif*. 2002; 26(1): 103–116.
11. Lorig KR, Ritter PL, Laurent DD, Plant K. Internet-based chronic disease self-management: A randomized trial. *Med Care*. 2006; 44(11): 964–971.
12. Ljotsson B, Lundin C, Mitsell K, et al. Remote treatment of bulimia nervosa and binge eating disorder: A randomized trial of Internet-assisted cognitive behavioural therapy. *Behav Res Ther*. 2007; 45(4): 649–661.
13. Strecher VJ, Shiffman S, West R. Randomized controlled trial of a web-based computer-tailored smoking cessation program as a supplement to nicotine patch therapy. *Addiction*. 2005; 100(5): 682–688.
14. Tate DF, Wing RR, Winett RA. Using Internet technology to deliver a behavioral weight loss program. *JAMA*. 2001; 285(9): 1172–1177.
15. Bond GE, Burr R, Wolf FM, et al. Preliminary findings of the effects of comorbidities on a web-based intervention on self-reported blood sugar readings among adults age 60 and older with diabetes. *Telemed J E Health*. 2006; 12(6): 707–710.
16. Gustafson DH, Hawkins R, Boberg E, et al. Impact of a patient-centered, computer-based health information/support system. *Am J Prev Med*. 1999; 16(1): 1–9.
17. Gerber BS, Solomon MC, Shaffer TL, Quinn MT, Lipton RB. Evaluation of an internet diabetes self-management training program for adolescents and young adults. *Diabetes Technol Ther*. 2007; 9(1): 60–67.
18. Glasgow RE, Davidson KW, Dobkin PL, Ockene J, Spring B. Practical behavioral trials to advance evidence-based behavioral medicine. *Ann Behav Med*. 2006; 31(1): 5–13.
19. Wangberg SC. An Internet-based diabetes self-care intervention tailored to self-efficacy. *Health Educ Res*. 2008; 23(1): 170–179.
20. Tate DF, Jackvony EH, Wing RR. Effects of Internet behavioral counseling on weight loss in adults at risk for type 2 diabetes: A randomized trial. *JAMA*. 2003; 289(14): 1833–1836.
21. Thompson D, Baranowski T, Cullen K, et al. Food, Fun and Fitness Internet program for girls: Influencing log-on rate. *Health Educ Res*. 2008; 23(2): 228–237.
22. Clarke G, Eubanks D, Reid E, et al. Overcoming Depression on the Internet (ODIN) (2): A randomized trial of a self-help depression skills program with reminders. *J Med Internet Res*. 2005; 7(2): e16.
23. Lenert L, Muñoz RF, Perez JE, Bansod A. Automated e-mail messaging as a tool for improving quit rates in an Internet smoking cessation intervention. *J Am Med Inform Assoc*. 2004; 11(4): 235–240.
24. Collins LM, Murphy SA, Strecher V. The Multiphase Optimization Strategy (MOST) and the Sequential Multiple Assignment Randomized Trial (SMART) new methods for more potent ehealth interventions. *Am J Prev Med*. 2007; 32(5 Suppl): S112–S118.
25. Strecher VJ, McClure JB, Alexander GL, et al. Web-based smoking-cessation programs: Results of a randomized trial. *Am J Prev Med*. 2008; 34(5): 373–381.
26. Norman GJ. Answering the “what works?” Question in health behavior change. *Am J Prev Med*. 2008; 34(5): 449–450.
27. Street LL, Luoma JB. Control groups in psychosocial intervention research: Ethical and methodological issues. *Ethics Behav*. 2002; 12(1): 1–30.
28. Mangunkusumo R, Brug J, Duisterhout J, de Koning H, Raat H. Feasibility, acceptability, and quality of Internet-administered adolescent health promotion in a preventive-care setting. *Health Educ Res*. 2007; 22(1): 1–13.
29. Celio AA, Winzelberg AJ, Wilfley DE, et al. Reducing risk factors for eating disorders: Comparison of an Internet- and a classroom-delivered psychoeducational program. *J Consult Clin Psychol*. 2000; 68(4): 650–657.
30. Patten CA, Croghan IT, Meis TM, et al. Randomized clinical trial of an Internet-based versus brief office intervention for adolescent smoking cessation. *Patient Educ Couns*. 2006; 64: 249–258.
31. Steele RM, Mummery WK, Dwyer T. Examination of program exposure across intervention delivery modes: Face-to-face versus internet. *Int J Behav Nutr Phys Activ*. 2007; 4:7.
32. Carlbring P, Nilsson-Ihrfelt E, Waara J, et al. Treatment of panic disorder: Live therapy vs. self-help via the Internet. *Behav Res Ther*. 2005; 43(10): 1321–1333.
33. Klein B, Richards JC, Austin DW. Efficacy of Internet therapy for panic disorder. *J Behav Ther Exp Psychiatry*. 2005; 37: 213–238.
34. Orbach G, Lindsay S, Grey S. A randomised placebo-controlled trial of a self-help Internet-based intervention for test anxiety. *Behav Res Ther*. 2006; 45: 483–496.
35. Rothert K, Strecher VJ, Doyle LA, et al. Web-based weight management programs in an integrated health care setting: A randomized, controlled trial. *Obesity*. 2006; 14(2): 266–272.
36. Severson HH, Gordon JS, Danaher BG, Akers L. ChewFree.com: Evaluation of a Web-based cessation program for smokeless tobacco users. *Nicotine Tob Res*. 2008; 10(2): 381–391.
37. Danaher BG, Smolkowski K, Seeley JR, Severson HH. Mediators of a successful Web-based smokeless tobacco cessation program. *Addiction*. 2008; 103: 1706–1712.
38. Ström L, Pettersson R, Andersson G. A controlled trial of self-help treatment of recurrent headache conducted via the Internet. *J Consult Clin Psychol*. 2000; 68(4): 722–727.
39. Spittaels H, De Bourdeaudhuij I, Vandelanotte C. Evaluation of a website-delivered computer-tailored intervention for increasing physical activity in the general population. *Prev Med*. 2007; 44(3): 209–217.
40. Jones M, Luce KH, Osborne MI, et al. Randomized, controlled trial of an internet-facilitated intervention for reducing binge

- eating and overweight in adolescents. *Pediatrics*. 2008; 121(3): 453–462.
41. Borkovec TD, Miranda J. Between-group psychotherapy outcome research and basic science. *J Clin Psychol*. 1999; 55(2): 147–158.
 42. Hursey KG, Rains JC, Penzien DB, Nash JM, Nicholson RA. Behavioral headache research: Methodologic considerations and research design alternatives. *Headache*. 2005; 45(5): 466–478.
 43. Koo M, Skinner H. Challenges of internet recruitment: A case study with disappointing results. *J Med Internet Res*. 2005; 7(1): e6.
 44. Watson JM, Torgerson DJ. Increasing recruitment to randomised trials: A review of randomised controlled trials. *BMC Med Res Methodol*. 2006; 6: 34.
 45. Gordon JS, Akers L, Severson HH, Danaher BG, Boles SM. Successful participant recruitment strategies for an online smokeless tobacco cessation program. *Nicotine Tob Res*. 2006; 8: S35–S41.
 46. McKay HG, Danaher BG, Seeley JR, Lichtenstein E, Gau JM. Comparing two Web-based smoking cessation programs: Randomized controlled trial. *J Med Internet Res*. 2008; 10(5): e40.
 47. Glasgow RE, Nelson CC, Kearney KA, et al. Reach, engagement, and retention in an Internet-based weight loss program in a multi-site randomized controlled trial. *J Med Internet Res*. 2007; 9(2): e11.
 48. Horrigan JB, Rainie L, Allen K et al. *The ever-shifting Internet population: A new look at Internet access and the digital divide*. Pew Internet & American Life, 2003. Retrieved July 23, 2009 from http://www.pewinternet.org/pdfs/PIP_Shifting_Net_Pop_Report.pdf
 49. Nielsen J. *Digital divide: The three stages*. useit com, 2006. Retrieved July 22, 2009 from <http://www.useit.com/alertbox/digital-divide.html>
 50. Kalichman SC, Weinhardt L, Benotsch E, Cherry C. Closing the digital divide in HIV/AIDS care: Development of a theory-based intervention to increase Internet access. *AIDS Care*. 2002; 14(4): 523–537.
 51. Kotler P, Roberto N, Lee NR. *Social Marketing: Improving the Quality of Life*. Thousand Oaks: Sage; 2002.
 52. Christensen H, Mackinnon A. The law of attrition revisited. *J Med Internet Res*. 2006; 8(3): e20.
 53. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: Toward a unified view. *MIS Quarterly*. 2003; 27(3): 425–478.
 54. Porter CS, Donthu N. Using the Technology Acceptance Model to explain how attitudes determine Internet usage: The role of perceived access barriers and demographics. *J Bus Res*. 2006; 59(9): 999–1007.
 55. Ritterband LM, Thorndike F. Internet interventions or patient education web sites? *J Med Internet Res*. 2006; 8(3): e18.
 56. Stanford J, Tauber ER, Fogg BJ, Marable L. *Experts vs. online consumers: A comparative credibility study of health and finance Web sites*. Consumer Reports Webwatch, 2002. Retrieved July 21, 2009 from <http://www.consumerwebwatch.org/pdfs/expert-vs-online-consumers.pdf>
 57. Graham AL, Bock BC, Cobb NK, Niaura R, Abrams DB. Characteristics of smokers reached and recruited to an internet smoking cessation trial: A case of denominators. *Nicotine Tob Res*. 2006; 8(Suppl 1): S43–S48.
 58. Alexander GL, Divine GW, Couper MP, et al. Effect of incentives and mailing features on online health program enrollment. *Am J Prev Med*. 2008; 34(5): 382–388.
 59. Danaher BG, McKay HG, Seeley JR. The information architecture of behavior change websites. *J Med Internet Res*. 2005; 7(2): e12.
 60. Eysenbach G. The law of attrition. *J Med Internet Res*. 2005; 7(1): e11.
 61. Frenn M, Malin S, Villarruel AM, et al. Determinants of physical activity and low-fat diet among low income African American and Hispanic middle school students. *Public Health Nurs*. 2005; 22(2): 89–97.
 62. Wikipedia: *Web analytics*. Wikipedia, 2008. Retrieved July 21, 2009 from http://en.wikipedia.org/wiki/Web_analytics
 63. Danaher BG, Boles SB, Akers L, Gordon JS, Severson HH. Defining participant exposure measures in Web-based health behavior change programs. *J Med Internet Res*. 2006; 8(3): e15.
 64. Associated Press. *Nielsen revises its gauge of Web page rankings*. Associated Press, 2008. Retrieved July 22, 2009 from <http://www.nytimes.com/2007/07/10/business/media/10online.html?ex=1341720000&en=7ad7f7ae739cb786&ei=5088&partner=rssnyt&emc=rss>
 65. Cobb NK, Graham AL, Bock BC, Papandonatos G, Abrams DB. Initial evaluation of a real-world Internet smoking cessation system. *Nicotine Tob Res*. 2005; 7(2): 207–216.
 66. McCoy MR, Couch D, Duncan ND, Lynch GS. Evaluating an Internet weight loss program for diabetes prevention. *Health Promot Int*. 2005; 20: 221–228.
 67. Danaher PJ, Mullarkey GW, Essegai S. Factors affecting website visit duration: A cross-domain analysis. *J Mark Res*. 2006; 43(2): 182–194.
 68. Danaher BG, Jazdzewski SA, McKay HG, Hudson CR. Bandwidth constraints to using video and other rich media in behavior change websites. *J Med Internet Res*. 2005; 7(4): e49.
 69. Peterson ET. *Web Site Measurement Hacks*. Sebastopol: O'Reilly Media; 2005.
 70. Williamson DA, Martin PD, White MA, et al. Efficacy of an internet-based behavioral weight loss program for overweight adolescent African-American girls. *Eat Weight Disord*. 2005; 10(3): 193–203.
 71. comScore: *Media matrix*. Comscore, 2009. Retrieved July 23, 2009 from <http://www.comscore.com/matrix/>
 72. Farvolden P, Denisoff E, Selby P, Bagby RM, Rudy L. Usage and longitudinal effectiveness of a Web-based self-help cognitive behavioral therapy program for panic disorder. *J Med Internet Res*. 2005; 7(1): e7.
 73. Andersson G, Carlbring P, Holmstrom A, et al. Internet-based self-help with therapist feedback and in vivo group exposure for social phobia: A randomized controlled trial. *J Consult Clin Psychol*. 2006; 74(4): 677–686.
 74. Hurling R, Catt M, Boni MD, et al. Using internet and mobile phone technology to deliver an automated physical activity program: Randomized controlled trial. *J Med Internet Res*. 2007; 9(2): e7.
 75. Brendryen H, Kraft P. Happy ending: A randomized controlled trial of a digital multi-media smoking cessation intervention. *Addiction*. 2008; 103(3): 478–484.
 76. Reynolds B, Dallery J, Shroff P, Patak M, Leraas K. A Web-based contingency management program with adolescent smokers. *J Appl Behav Anal*. 2008; 41(4): 597–601.
 77. Strecher VJ, Marcus A, Bishop K, et al. A randomized controlled trial of multiple tailored messages for smoking cessation among callers to the cancer information service. *J Health Commun*. 2005; 10(Suppl 1): 105–118.
 78. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behav Res Methods*. 2008; 40(3): 879–891.
 79. Christensen H, Griffiths K, Groves C, Korten A. Free range users and one hit wonders: Community users of an Internet-based cognitive behaviour therapy program. *Aust NZ J Psychiatry*. 2006; 40(1): 59–62.

80. Lenert L. *Automating smoking cessation on the web*. WATI: Web-assisted Tobacco Intervention, 2005. Retrieved July 22, 2009 from <http://www.wati.net/presentations/watiII.lenert.pres.20050607.pdf>
81. Strecher VJ, Shiffman S, West R. Moderators and mediators of a Web-based computer-tailored smoking cessation program among nicotine patch users. *Nicotine Tob Res.* 2006; 8(S1): S95–S101.
82. Tullis T, Albert W. *Measuring the User Experience: Collecting, Analyzing, and Presenting Usability Metrics*. San Francisco: Morgan Kaufmann; 2008.
83. Tullis TS, Stetson JN. *A comparison of questionnaires for assessing website usability*. Usability Professionals Association Conference, Minneapolis, MN, June 7–11, 2004. Retrieved July 22, 2009 from <http://home.comcast.net/%7Etomtullis/publications/UPA2004TullisStetson.pdf>