Available online at www.sciencedirect.com
SCIENCE DIRECT®
Journal of Operations Management 22 (2004) 1–21
Journal of Operations Management
www.elsevier.com/locate/dsw

New measurement scales for evaluating perceptions of the technology-mediated customer service experience

Craig M. Froehle[a,*], Aleda V. Roth[b,1]

[a] University of Cincinnati, College of Business, 521 Lindner, P.O. Box 210130, Cincinnati, OH 45221-0130, USA
[b] The Kenan-Flagler Business School, The University of North Carolina at Chapel Hill, Campus Box 3490, McColl Hall, Chapel Hill, NC 27599-3490, USA

Received 31 December 2002; received in revised form 13 October 2003; accepted 1 December 2003

Abstract

Service organizations are increasingly utilizing advanced information and communication technologies, such as the Internet, in hopes of improving the efficiency, cost-effectiveness, and/or quality of their customer-facing operations. More of the contact a customer has with the firm is likely to be with the back-office and, therefore, mediated by technology. While previous operations management research has been important for its contributions to our understanding of customer contact in *face-to-face* settings, considerably less work has been done to improve our understanding of customer contact in what we refer to as technology-mediated settings (e.g., via telephone, instant messaging (IM), or email). This paper builds upon the service operations management (SOM) literature on customer contact by theoretically defining and empirically developing new multi-item measurement scales specifically designed for assessing *technology-mediated* customer contact. Seminal works on customer contact theory and its empirical measurement are employed to provide a foundation for extending these concepts to technology-mediated contexts. We also draw upon other important frameworks, including the Service Profit Chain, the Theory of Planned Behavior, and the concept of media/information richness, in order to identify and define our constructs. We follow a rigorous empirical scale development process to create parsimonious sets of survey items that exhibit satisfactory levels of reliability and validity to be useful in advancing SOM empirical research in the emerging Internet-enabled back-office.

Keywords: Service operations; Measurement and methodology; Empirical research; Technology

1. Introduction

Service organizations are increasingly utilizing advanced information and communication technologies, such as the Internet, in hopes of improving the efficiency, cost-effectiveness, and/or quality of their customer-facing operations (Fitzsimmons and Fitzsimmons, 2004; Huete and Roth, 1988; Haynes and Thies, 1991; Hill et al., 2002; Oliveira et al., 2002; Roth, 2000; Boyer et al., 2002). More of the contact a customer has with the firm is likely to be with the back-office and mediated by technology (e.g., via telephone, instant messaging (IM), or email). Service organizations are finding that interacting with their customers via these new technologies can be a significant challenge (Zeithaml et al., 2002). While Internet-based customer support can potentially reduce costs on a per-transaction basis, customer

* Corresponding author. Tel.: +1-513-556-7174; fax: +1-513-556-5499.
E-mail addresses: craig.froehle@uc.edu (C.M. Froehle), aleda_roth@unc.edu (A.V. Roth).
[1] Tel.: +1-919-962-3181; fax: +1-919-962-4266.

0272-6963/$ – see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.jom.2003.12.004

satisfaction, as well as long-term customer loyalty, can be severely damaged by a bad on-line experience (Barnes et al., 2000; Heim and Sinha, 2001a). Therefore, the service operations management (SOM) issues related to managing customer contact in these technology-mediated environments warrant attention.

The concept of customer contact (Chase, 1978, 1981) has long been an integral element of SOM research. Customer contact is generally held to be a function of the interaction between a customer and a service provider (Kellogg and Chase, 1995). Initial work to empirically define the underlying dimensions of customer contact was performed using a hospital setting, an environment in which all contact between customers and employees occurred in-person and face-to-face (Kellogg and Chase, 1995; Soteriou and Chase, 1998). While extremely valuable, it is unclear whether the results of this research are equally applicable to contexts involving customer contact in technology-mediated (i.e., non-face-to-face) service delivery processes. We extend the abilities of SOM researchers to examine these environments by adapting Chase's initial ideas of customer contact to these new technology-mediated contexts, using customer perceptions.

This paper builds upon prior literature by theoretically defining the conceptual domains of inquiry, constructs, and operational measures specific to advancing SOM research in technology-mediated customer contact situations. This research context is particularly applicable for the emerging area of e-services. The unit of analysis in this research is the customer.

We follow a normative two-step process. First, we identify ten theory-based constructs covering three domains that comprise antecedents and consequences of technology-mediated customer contact from a customer's perspective. Second, because the constructs are latent (i.e. non-observed) variables, we apply a rigorous procedure for ensuring the psychometric adequacy of the resulting new multi-item measurement scales.

While Kellogg and Chase's (1995) seminal work on customer contact theory and the hypothesized Service Profit Chain (Heskett et al., 1994) helps motivate and structure our thinking in the SOM arena, we also draw upon the Theory of Planned Behavior (Ajzen, 1985, 1991), and incorporate the concept of media/information richness (Daft and Lengel, 1984, 1986) in identifying and defining our constructs. Once the constructs are defined, we then follow a rigorous empirical scale development process in order to identify parsimonious sets of survey items that exhibit satisfactory levels of reliability and validity.

Section 2 presents a brief background of the research context and defines and illustrates the specific constructs for which new measurement scales are developed. The third section provides details on the preliminary scale development methodology and field study database. Section 4 describes and reports on confirmatory modeling results. In Section 5, we conclude with a discussion of the implications of our results and usage of the scales, review the limitations of our study, and offer some concluding thoughts.

2. Background

The advancement of communications and information technology is altering the ways that customers interface with service providers and, therefore, may influence customers' perceptions of the service experience (Boyer et al., 2002; Heim and Sinha, 2001b; Huete and Roth, 1988; Oliveira et al., 2002; Roth, 2000, 2001). This paper specifically focuses on a particular context that we call *technology-mediated customer contact*. This context is one of five possible modes (or distinct conceptual archetypes) of customer contact in relation to technology (see Fig. 1).

The first archetype, depicted in Fig. 1A, is called *technology-free customer contact.* The customer is in physical proximity of, and interacts with, a human service provider (hereafter, referred to as "service rep"). This archetype typifies traditional notions of face-to-face service encounters emphasized by Chase (1978). In technology-free customer contact, technology per se does not play a direct role in providing the service. Examples of technology-free customer contact include a psychiatrist's in-office consultation with a patient, a retail bank teller exchanging a customer's coins for paper currency, or an old-fashioned, general store clerk transacting the sale of merchandise with a cash drawer.

The second archetype is *technology assisted-customer contact.* The service rep employs technology as an aid to improve the face-to-face contact, but

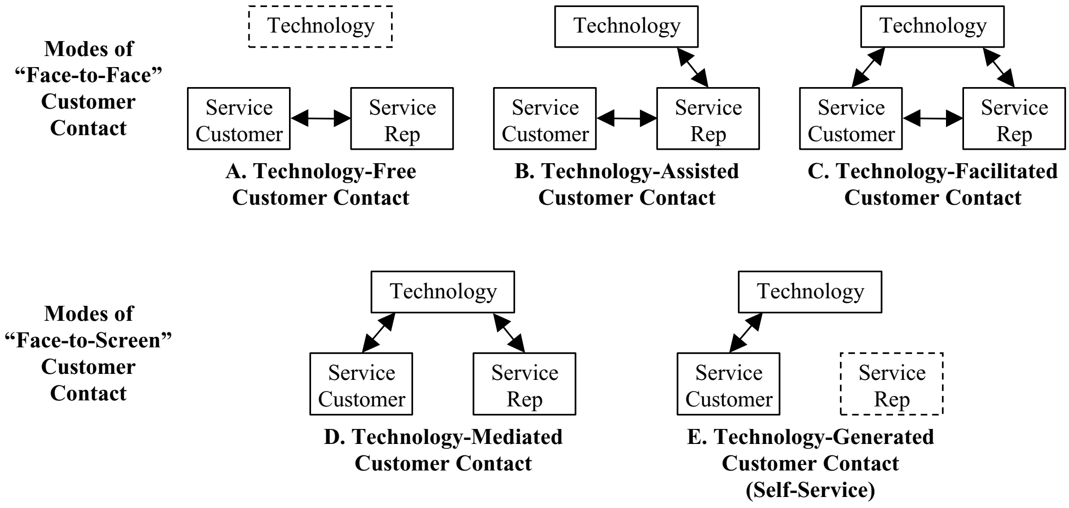

Fig. 1. Conceptual archetypes of customer contact in relation to technology.

the customer does not have access to the technology (Fig. 1B). This situation often occurs during airline check-in; the service representative interacts with a computer terminal, but the customer does not.

Third, we refer to *technology-facilitated customer contact* where, during the face-to-face service encounter, both the service representative and the customer have access to the same technology (Fig. 1C). Here technology is employed to enhance the face-to-face communication between a customer and a service provider, such as when a financial consultant uses PowerPoint in a meeting with a client.

Fourth, as depicted in Fig. 1D there is *technology-mediated customer contact*, where the customer and the human service representative are not physically co-located. Therefore, the service encounter is not a traditional face-to-face contact. To enable communication, some form of technology must be employed, such as when a voice telephone call or online instant messaging is initiated with a customer service rep in a back-office call center.

Finally, there is *technology-generated customer contact*, where the human customer service representative component of the service encounter is entirely replaced by technology (Fig. 1E). This is the most technology-intensive situation. For example, bank ATMs, self-service kiosks, automated car washes, and website-based knowledge-bases offer the option of service without the assistance of human service reps. Self-service technology has been subjected to SOM scholarly inquiry for more than two decades (Sasser et al., 1978; Collier, 1983).

In anticipation of technological progress, the latter two modes of customer contact (Fig. 1D and E) can be generally referred to as "face-to-screen" archetypes. Even though the technology being used for the customer contact may not explicitly employ an actual display *screen* (e.g. a telephone), we speculate that as technologies become more digitized, customers will increasingly use some form of display, such as ATM screens or picture phones, to communicate with the service provider.

As indicated earlier, technology-mediated customer contact (Fig. 1D), the focus of this paper, is not well understood, and the literature is void of constructs and measurement scales for online customer contact. Developing robust and valid measurement scales is the first step in advancing SOM theory-building and testing in this area.

2.1. Conceptual domains

Normative conceptual models in SOM, such as the Service Profit Chain (Heskett et al., 1994), hypothesize that the profitability and growth of a service firm depend on its ability to create loyal customers. However, the specific antecedents of customer loyalty have not yet been generally agreed upon in the

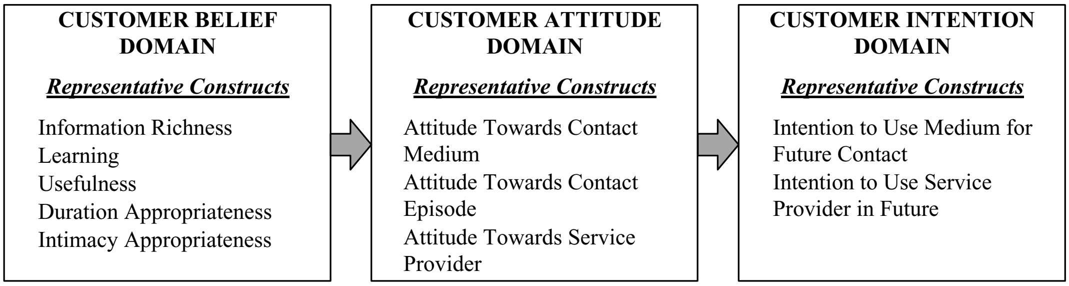

Fig. 2. Conceptual B–A–I framework of technology-mediated customer service—hypothesized domains and constructs.

context of on-line customer support services. We posit that both psychological (or unobserved) factors and physical properties associated with use of a technology medium will be important for successful on-line customer contacts. Being able to eventually determine these psychological antecedents requires us first to develop robust means for defining and measuring them. Therefore, in this section, we use several theoretical bases in order to guide our construct selection and operational definitions that are applicable for SOM.

Drawing from the social sciences literature, two of the more influential and empirically supported theories explaining individual and customer actions are the Theory of Planned Behavior (Ajzen, 1985, 1991) and its predecessor, the Theory of Reasoned Action (Fishbein and Azjen, 1975). Those models posit that a customer's beliefs, or rational cognitive assessments, of the contact episode[2] influence various customer attitudes towards that episode. Fishbein and Azjen (1975) describe attitude as "a person's general feelings of favorableness or unfavorableness towards some stimulus object" (p. 6) and is conceptually similar to what business academics often refer to as "satisfaction." These attitudes, in turn, drive customer intentions for future behavior, such as loyalty to a service provider (as suggested by the Service Profit Chain and customer contact theory). This belief → attitude → intention (B–A–I) model lies at the heart of many contemporary management theories (e.g., the Technology Acceptance Model, Davis et al., 1989).

In this paper, we applied the general structure of the B–A–I model, as shown in Fig. 2, which suggests that potentially important psychological constructs will fall into three broad domains: beliefs, attitudes, and intentions. The specific constructs in each of the three B–A–I categories of Fig. 2 were identified by reviewing previous research as well as discussions with managers responsible for their firms' technology-mediated customer service functions (these were typically characterized as "back-office" support services).

This grounded theory approach guides our thinking for developing the measurement scales we require in order to assess technology-mediated customer service. Note also the primary purpose of this paper is to define a set of theoretically grounded psychological constructs and to devise reliable and valid operational measures for each. The measurement task of constructing ten new multi-item measurement scales of psychological phenomena over three conceptual domains is sufficiently complex that it warrants a separate paper. We leave it to future research to test specific hypothesized paths among these constructs. Each of the ten constructs shown in Fig. 2 is defined and described more extensively in the subsections that follow. In Section 3, we present the details of our research approach.

2.2. Belief constructs

In our conceptual framework, five constructs are covered in the belief domain. This belief domain constitutes the hypothesized first-order antecedents

[2] In this study, a 'contact episode' is defined as an exchange of communication between a customer and a service representative. This communication exchange may occur over one or several different service encounters with one or more service representatives, a situation sometimes referred to as a "pseudo-relationship" (Ghutek et al., 1999).

of attitudes, which, in turn, lead to customers' intentions.

2.2.1. Information richness belief (IRB)

The *information richness belief* construct taps the customer's cognitive assessment of the complexity and vividness of the communication between the provider and the customer, parallel to the original description of Media Richness offered by Daft and Lengel (1984, 1986). That original research posited a hierarchy among communication media, with face-to-face communication modes being the richest and all other media being less rich (see Fig. 1 (p. 196) of Daft and Lengel, 1984). Therefore, it seems reasonable that one could assess the media richness of a technology-based medium by assessing how closely it approximates the richness of face-to-face communication experience.

Kellogg and Chase (1995) define Information Richness as a multidimensional construct calculated as the average of four variables: feedback speed, the type(s) of channels used, the major topic of discussion, and the kind of language used. While there is still some debate over the relative merits of multidimensional constructs in management research, some critics contend that they are "conceptually ambiguous" (Edwards, 2001) and should therefore be avoided whenever possible. Furthermore, as information is generally viewed as an experience-level phenomenon (Carlson and Zmud, 1999), it is consistent with previous literature to measure information richness as an experience-level construct. Due to concerns regarding measurement validity, as well as inconsistent support in the literature for the individual dimensions proposed by Kellogg and Chase, we develop a construct-level scale for measuring Information Richness beliefs.

2.2.2. Learning belief (LB)

A customer's *learning belief* is defined as the belief that he/she increased his/her own knowledge, capacity for understanding, or perspective-taking during the contact episode (Boland and Tenkasi, 1995). Learning has often been considered an antecedent for usefulness, especially if that is a goal of the customer or motivation for his/her presence in the learning situation (Alavi, 1994; Alavi et al., 1997). Moreover, learning appears important to e-service processes as well, since the information content of many e-service product offerings is substantial (Porter and Millar, 1985; Oliveira et al., 2002). Thus, it seems potentially valuable to be able to measure the LB construct in a technology-mediated customer service context.

2.2.3. Usefulness belief (UB)

Our inclusion of the *usefulness belief* extends previous definitions of customer contact by also considering the degree to which the contact episode fulfills the customer's perceived needs and desires. Generally, customers will be more motivated to use a service again when they benefit or derive value from it (Walley and Amin, 1994; Heim and Sinha, 2001b) and, therefore, perceive it as 'useful.' Also, usefulness was determined to be a vital element in the Technology Acceptance Model (Davis et al., 1989), giving precedence for its consideration here. Finally, the UB is an elemental component of determining the customer-perceived value (or utility) of the service to the customer. This is the perceived value from the tangible and intangible benefits, relative to costs, that a customer receives from the service bundle (Heskett et al., 1994, 2003; Roth and Menor, 2004).

2.2.4. Duration appropriateness belief (DAB)

This construct represents the customer's belief about the duration of the contact episode. Kellogg and Chase (1995) found it to be their most influential determinant of customer contact. Later, Soteriou and Chase (1998) refined the definition of this construct and empirically demonstrated that customers perceive a certain *range* of contact time as most preferable for a contact episode. Durations shorter or longer than that most preferred range lead to decreased perceptions of service quality. It seems likely that a similarly non-linear relationship between contact time and service satisfaction should hold here as well, thus inciting us to include duration, or more specifically, duration *appropriateness*, as a key customer contact belief.

It is important to note that our definition of DAB considers only the time the customer is *actively* involved in communication with the service provider's customer service function, and does not explicitly (or implicitly) attempt to differentiate, or separate out, the time the customer initially spends waiting (in queue or "on hold") from the period of *active* communication with the service provider (e.g., talking or listening on the telephone, reading or composing an email or instant message, etc.). While there may be motivation from other research streams to incorporate this distinction in future extensions of our research, the customer contact literature currently focuses on *active* contact time, and we maintain that distinction in our working definition of this construct.

2.2.5. Intimacy appropriateness belief (IAB)

The *intimacy appropriateness belief* relates to the level of "mutual confiding and trust" established during the customer contact episode, and is among the key customer contact variables discovered by Kellogg and Chase (1995). Intimacy has been further validated in both the academic and practitioner literatures (Pine and Gilmore, 1998), and is seen as a crucial element in meaningful contact, both with customers and employees (DeSanctis and Monge, 1999). Like the duration belief, however, more intimacy in a customer-provider context is not always better, and customers have preferred levels of intimacy (Soteriou and Chase, 1998). In the context of marketing, there is precedence for considering intimacy, as it is highly related to the *empathy* dimension in the SERVQUAL instrument (Parasuraman et al., 1988).

2.3. Attitudinal constructs

Three attitudinal constructs in our framework differ by the customer's perception of the stimulus, or object, of the attitude (e.g., the communications technology, the overall service experience, and the service provider).

2.3.1. Attitude towards the contact medium (AM)

Attitude towards the contact medium reflects the immediate favorableness (or customer-perceived satisfaction) or unfavorableness (dissatisfaction) created by using the communication medium (e.g., telephone, email or instant messaging) employed during the contact episode. The target of the customer's affect is limited to the technology-based communication medium used in the contact episode.

2.3.2. Attitude towards the contact episode (AE)

Attitude towards contact episode reflects the immediate favorableness (customer-perceived satisfaction) or unfavorableness (dissatisfaction) with the service content of the entire contact episode. In contrast to the first attitude construct above, which reflects satisfaction with only the technology-related process element, this AE construct reflects the customer's overall attitude towards the entire customer contact episode (i.e., the total customer service experience). This construct also differs from the previous attitude construct in that the "stimulus object" is the contact service *episode* rather than the communication *medium* employed.

2.3.3. Attitude towards the service provider (AP)

This attitude reflects the level of general satisfaction or dissatisfaction the customer has with the service provider, or the company from whom the customer is currently seeking service, at the end of the contact episode. We recognize that, in virtually all cases, the customer's attitude toward the service provider will likely be influenced by previous contact episodes, moments of truth, and other experiences, and will not be entirely based on only a single technology-mediated customer contact episode. However, since every contact a service provider has with a customer may influence that customer's attitude towards the service provider, it is important to include this construct in the model shown in Fig. 2. The "stimulus object" for this attitude construct is the service provider, which differs from the two attitude constructs described previously.

2.4. Intention constructs

The two constructs composing the Intention domain are proxy variables for customer loyalty.

2.4.1. Intention to use medium again (IM)

The *intention to use the medium again* construct assesses the customer's internally estimated likelihood that he will carry out future contact with the service provider via the same communications medium as was used for his most recent contact episode. The user's *intention* to adopt and use a specific technology, such as email or instant messaging, has been shown to be a valuable predictor of actual future technology use (Davis et al., 1989), and is therefore a valuable construct to include here. Consistent with the first six constructs presented, the object of this intention construct is the communication medium.

2.4.2. *Intention to use service provider again (IP)*

This construct assesses the customer's internally estimated likelihood that he will make use of the firm's services again at some future time (or continue the relationship if ongoing). This construct provides some measure of the probability that a customer will remain loyal to the service provider (Heskett et al., 1994; Wiersema, 1996; Reichheld and Schefter, 2000), as it involves a direct linkage between decisions and results. While satisfaction (attitude) is a highly important measure in its own right, the intentions of the customer (i.e., customer loyalty), when taking that satisfaction into account, are significant drivers of the service provider's profitability and long-term viability (Pontes and Kelly, 2000; Coffee, 2001). Consistent with the construct presented in Section 2.4.1, the object of this intention construct is the service provider.

3. New scale development methods

Scale development, or the iterative design and refinement of multi-item scales employed to measure the constructs we are interested in studying, is vital to empirical research in operations management (Adam and Swamidass, 1989; Flynn et al., 1990; Hensley, 1999; Menor, 2000; Stratman and Roth, 2002; Roth and Schroeder, forthcoming). Without reliable and valid measurement scales, we are less capable of empirically estimating the relationships that tie various operational concepts together. Establishing the *validity* of the scales is dependent first upon establishing that they are *reliable* measures (Churchill, 1995). Fig. 3 illustrates the general methodological steps involved in the scale development process we followed. The goal of our process is to locate or create reliable and valid multi-item scales for measuring the ten constructs described in Section 2. The content validity of these constructs was tentatively established by extensive literature reviews and interviews with managers and customers of technology-mediated services. This section describes how the items and multi-item scales were developed following the steps outlined in Fig. 3. Appendix A provides an illustrative example of how items were refined over extensive testing and Appendix B shows the final set of items retained.

3.1. *Item generation*

The construct definitions provided in Section 2 are necessary, but not sufficient, to advance our understanding of the managerial and operational relationships that shape the technology-mediated customer service experience. Thus, the first step in constructing new multi-item measurement scales is to generate sets of items that tap into the latent constructs and permit us to accurately and reliably assess these constructs from customers' perspectives (Churchill, 1995). Some of the constructs involved in this research have been operationalized in previous studies and scales were available for these constructs. However, none of the existing scales was exactly appropriate for re-application in the context of technology-mediated services. Table 1 shows those constructs for which

Table 1
Measurement Scales Employed in Prior Research

Construct	Reference	Description of scale	Limitations to re-application
Duration	Kellogg and Chase (1995)	Objective, absolute measurement (i.e. seconds)	Per Soteriou and Chase (1998), objective time durations do not correspond exactly to duration *appropriateness* (our construct)
Duration	Soteriou and Chase (1998)	Objective, absolute measurement (i.e. seconds)	Per Soteriou and Chase (1998), objective time durations do not correspond exactly to duration *appropriateness* (our construct)
Intimacy	Soteriou and Chase (1998)	Three-item Likert scales	Wording is particular to the context (hospital setting) of the original study
Information Richness	Kellogg and Chase (1995)	Used four-item, averaged, multi-dimensional scale	Scale did not demonstrate appropriate discriminant or convergent validity
Usefulness	Agarwal and Prasad (1999)	Eight-item scale measured “perceived usefulness” of a new technology	Highly reliable (α=0.95), but long (eight items) and context-specific (adoption of new technology)

Item Generation

Step 1 – Literature Review and Scale Creation
- Examine literature for existing scales
- Assess applicability of existing scales and revise if appropriate
- Develop new items and constructs as necessary

Item Refinement

Step 2 – Pre-Test
- Employ convenience sample to participate in written survey
- Receive qualitative feedback and evaluate item integrity
- Eliminate or reword problematic items and augment with new items as needed

Step 3 – Pilot Test #1
- Develop web-based survey instrument
- Test scales and items using convenience sample
- Calculate reliability scores and modify scales as needed

Step 4 – Pilot Test #2
- Test scales and items using representative random sample
- Calculate reliability scores and modify scales as necessary

Confirmatory Analysis

Step 5 – Gather Field Dataset
- Randomly sample service customers and send participation invitation emails
- Collect response data using web-based survey
- Randomly divide field dataset into calibration and holdout samples

Step 6 – Initial Confirmatory Factor Analysis (CFA) of Calibration Sample
- Perform CFA on initial measurement model
- Examine overall fit and inspect item-level fit for multi-dimensionality
- Remove "problem" indicators, if any, from measurement model

Step 7 – Final Confirmatory Factor Analysis (CFA)
- Assess modified scales for acceptable reliability using calibration sample
- Assess global fit and parameter fit using calibration sample
- If no changes required, assess global fit and parameter fit using holdout sample

Fig. 3. Methodological steps in the scale development and data gathering process.

scales already existed, their sources, and our comments.

As shown in Fig. 1, the technology-mediated customer contact archetype differs significantly from the traditional face-to-face contact archetype for which these prior scales were originally developed. Therefore, we found it necessary to construct entirely new scales for all ten constructs presented in Section 2. Where possible, the wording of initial items generated for each construct was adapted from existing scales and literature. Next we describe how these items were refined.

3.2. Iterative item refinement

Once the initial set of items for each construct was specified (through review of the literature and discussions with managers), pre-testing and pilot-testing took place to ensure that the initial measurement scales (and items) were as reliable and as valid as can be determined. Pre-testing and pilot testing are important steps in the scale development process. Anderson and Gerbing (1991) and Stratman and Roth (2002) demonstrated that performing a pre-test assessment was useful in culling out invalid measures (i.e. those unlikely to be supported in a confirmatory factor analysis).

3.2.1. Pre-testing

The first pre-test employed a convenience sample of 39 MBA students using a pencil-and-paper survey that assessed their reactions to a controlled, in-class technology-mediated communication session. The emphasis at this early stage was to enhance readability and clarity of the questions as well as to enhance the items' and emerging scales' content validity. Note the intention domain scales were not part of the scope of the effort at this pre-test stage. Because MBA students were Internet users and were familiar with the context, we asked them for qualitative feedback on the items. Based on their comments, we modified the items to improve both their reliability and their parsimony by (a) deleting troublesome items, (b) rewording items that showed promise but were confusing to the respondents, and (c) adding new items to "shore up" scales that had undesirable items removed (Appendix A shows an example of how the items composing a scale evolved over the several iterations of scale development).

3.2.2. Pilot tests

At this point, it was deemed desirable for pilot testing the instrument to obtain further data using electronic means, if possible, for we knew that the ultimate field application of the scales would be done using a web-based survey tool. Web-based surveys are the least expensive to produce and maintain, eliminate the possibility of missing data, and facilitate the rapid and accurate assembly of the responses into a complete dataset. Given the technology-mediated context of the study, another advantage to using a web-based medium for data collection is the fact that our respondents are likely to already be familiar with the Internet. Email-based surveys are an alternative approach, but are more costly to administer, more time-consuming to parse into a dataset, and have so far achieved disappointing response rates overall (Parzinger, 1999).

We conducted two pilot tests of the survey instrument. The first pilot test of these ten new multi-item measurement scales involved a convenience sample of MBA and undergraduate students. A web-based survey, which asked about their most recent online or telephone customer service encounter, was developed specifically for this study (and was used for both pilot tests). The pilot test respondents were recruited principally through email solicitation and through announcements made in classrooms. To motivate participation, an incentive of a free soft drink at the business school cafeteria was offered to all who completed the survey. The website remained available for one week, at which point we had obtained 35 responses. As can be seen in Table 2, all scales surpassed the minimally acceptable level of 0.70 (Nunnaly, 1978; Carmines and Zeller, 1979) with the exception of the Information Richness scale. Also, at six items, the Usefulness scale is rather unwieldy. For these reasons, a second pilot test was performed after we modified select items and scales.

It was deemed desirable for the second pilot test to employ a larger and more generalizable sample so that we could have more statistical power and, therefore, the greatest possible confidence in our scales prior to applying them in a full field study. For this reason, an arrangement was made with an Internet-based company to allow us access to its mailing list of users of its website. We sent emails to 3117 randomly

Table 2
Pilot test scale reliability scores and number of items per scale

Scale	Pilot test #1 ($n = 35$)	Pilot test #2 ($n = 305$)
Information richness (IRB)	0.67 (4)	0.83 (3)
Learning (LB)	0.75 (4)	0.89 (4)
Usefulness (UB)	0.92 (6)	0.91 (4)
Duration appropriateness (DB)	0.78 (2)	0.87 (2)
Intimacy appropriateness (IB)	0.83 (3)	0.82 (3)
Attitude towards medium (AM)	0.86 (3)	0.92 (3)
Attitude towards episode (AE)	0.89 (4)	0.94 (3)
Attitude towards provider (AP)	0.92 (4)	0.95 (3)
Intention towards medium (IM)	0.90 (3)	0.89 (3)
Intention towards provider (IP)	0.97 (3)	0.96 (3)

Scores shown are Cronbach's α (# of items in scale shown in parentheses).

selected users of the firm's website and received 305 completed web-based surveys for a response rate of 9.8%. In order to gain access to the mailing list, we promised complete anonymity to the company and its user-base, so no personal or demographic data were collected. The results of the second pilot test (Table 2) show that all scales surpass the 0.80 level for coefficient α, which signifies that the scales have a high degree of internal reliability (Nunnaly, 1978; Carmines and Zeller, 1979).

3.3. Customer database and field survey

Satisfied by the apparent reliability and parsimony of our new measurement scales, we moved into the next phase of testing our survey instrument in a field setting (step 5 in Fig. 3). For this phase, we obtained the sponsorship of well-known, established Internet service provider, whose market research area gave us access to its user base and to its expertise in survey research. As a condition of participating in the study, our corporate partner requested complete anonymity. Therefore, we refer to it here as Iotanet, and some descriptive figures have been modified to protect its identity.

Iotanet survey researchers indicated that our survey instrument was too long, and out of pragmatic concerns for ensuring a higher completion rate, adequate incentive levels and higher quality responses, we reduced our final survey length. For these reasons, we had targeted the total number of items to four or less per construct. Because of our extensive pre-testing and pilot work, we felt confident that the scales represented very parsimonious measures of the respective constructs and that we had an acceptable number of items for confirmatory modeling.

At the time the survey was implemented, Iotanet was one of the world's largest Internet service providers (ISP) with over 10 million subscribers. It had been a leader in providing multi-media customer service, offering support via voice telephone, email, and instant messaging (chat). Iotanet employed over 5000 full-time, part-time, and contract customer service representatives, and outsourced some elements of its back-office customer support function. Iotanet's contact centers included customer service representatives (CSR), or "consultants," dedicated to a particular medium (i.e. telephone, email and chat) for any given shift. Management of these contact centers and CSR teams is a primary responsibility of the operations management function at Iotanet.

At the time of the study, Iotanet's internal customer support database provided some background information about its customers, but it was neither exhaustive nor state-of-the-art compared to more sophisticated and expansive customer relationship management (CRM) systems. Because of Iotanet's pervasive customer base (a significant percentage of all Internet users were Iotanet subscribers), and given the magnitude of its on-line customer service operations (over 100,000 customer support requests are handled each day), Iotanet was seen as able to provide an ideal context for our survey. Over a 6-day period, 12,050 email invitations were sent out to randomly sampled customers who had contacted an Iotanet customer service representative in the previous 24 h. Within 24 hours of the last invitation being sent out, a total of 2001 surveys had been completed yielding a response rate of 16.6%.

One concern of sample surveys is the possibility of non-response bias (Churchill, 1995; Fowler, 1988), which is the potential that those members of the sampling frame who elected to participate in the study are somehow systematically different from those who chose not to. If present, non-response bias limits the generalizability of the findings, for the sample is not truly random and, therefore, may not be representative of the sampling frame or overall population. Non-response bias is a difficult issue to both detect and correct (if present), for the bias could manifest itself in varying degrees along any (or all) of the variables measured. Moreover, unless the demographic data are tightly linked to the phenomena under scrutiny, even testing differences in demographic statistics is unlikely to uncover any influential, systematic non-response bias.

In order to address the possibility of non-response bias, we requested overall descriptive statistics about Iotanet's customer base. Iotanet management was unable to provide this due to its customer privacy policy. As a contingency, we gathered representative demographic data for Internet users at large for the point in time of our study. These statistics are shown Table 3, which provides a rough comparison with similar statistics from our sample. Our sample is generally consistent with the population of Internet

Table 3
Comparative demographics of Iotanet sample and Internet population

	Iotanet sample (n = 2001) (%)	Internet population (%)
Gender[a]		
Men	45	52
Women	51	48
N/A[c]	4	–
Age[a]		
Under 25	10	14
26–64	75	79
Over 65	11	4
N/A[c]	4	3
Nationality[b]		
US/Canada	91	50
Latin America	1	3
Europe	2	29
Australasia	<1	17
Africa	<1	1
N/A[c]	5	–

[a] Source for Internet statistics: Nie and Erbing (2000).
[b] Source for Internet statistics: CommerceNet (2000).
[c] Providing these data was optional for our respondents.

users as a whole with the exception that it is somewhat more North America-centric. Iotanet is a North American company and a majority of its customers reside in North America. Also, customer surveys were available only in English.

One concern of all *self-report* studies is the potential existence of common method variance “contaminating” the data and disguising the “true” variance. Because the same individual provides responses for all our measures, any unusual variance in that respondent (i.e. variance outside of what would normally be expected) will be reflected in all measures (Podsakoff and Organ, 1986). Common method variance is present when correlations between measures are due to the fact that the same individual provides the responses for both measurement scales rather than any true relationship between the constructs. One approach for assessing whether this condition exists is Harmon’s one-factor test (Podsakoff and Organ, 1986). This approach employs a factor analysis of all relevant variables. If common method variance is present to a significant degree, a *single* factor will emerge (Podsakoff and Organ, 1986; Miceli et al., 1991). We performed an exploratory (principal components) factor analysis with no rotation and found that six factors had Eigenvalues exceeding one. While not conclusive proof that common method variance does not exist in our data, it suggests that any common method variance that *does* exist is unlikely to be problematic (Frohlich and Westbrook, 2002).

4. Confirmatory factor analysis (CFA)

The final steps in the scale development process of Fig. 3 were confirmatory factor analyses (CFA), or measurement modeling. In contrast to the more widely used, and generally atheoretical, exploratory factory analysis (EFA), CFA involves the estimation of an *a priori* measurement model, wherein the observed variables are mapped onto the latent constructs according to theory and prior testing by the researcher (Bollen, 1989). The primary benefit of CFA is the ability to test measurement scales for evidence of convergent and discriminant validity (Pedhazur and Schmelkin, 1991). This is primarily accomplished by examining the factor loadings of latent constructs on the indicators.

4.1. Initial measurement model and analysis

Fig. 4 graphically depicts the initial, hypothesized measurement model. For clarity, many latent constructs, observed indicators, covariances, and error terms are not shown. This measurement model represents the state of our scales at the conclusion of the second pilot test, and is represented mathematically in Eq. (3a) and (b) below (3a is the simplified form and 3b is the matrix form, both employing standard covariance modeling notation).

$$y = \Lambda_y \eta + \varepsilon \tag{3a}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{30} \\ y_{31} \end{bmatrix} = \begin{bmatrix} \lambda_1 & \lambda_2 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & \cdots & \lambda_{30} & \lambda_{31} \end{bmatrix} \begin{bmatrix} \eta_1 \\ \eta_2 \\ \vdots \\ \eta_9 \\ \eta_{10} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_{30} \\ \varepsilon_{31} \end{bmatrix} \tag{3b}$$

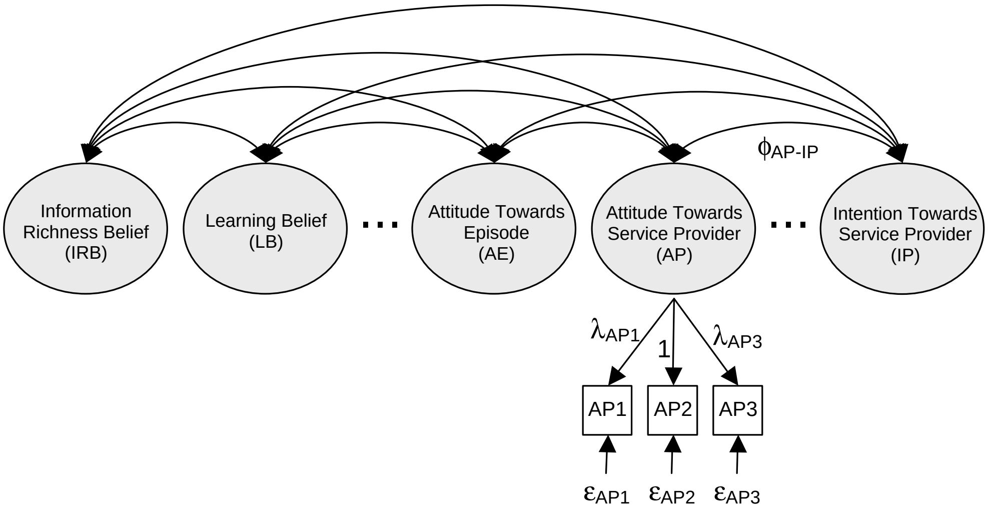

Fig. 4. Measurement Model—graphical representation. This diagram illustrates the measurement model structure involved in the confirmatory factor analyses (CFAs). Note that five of the ten latent constructs, most of the covariances (Φ), and all but three of the observed indicators and error terms, are not shown for clarity.

We confirmed that this model is *identified* using the "Two Indicator Rule" as outlined by Bollen (1989). In order to test this measurement model, we employed a holdout sample approach. In this approach, we randomly divided our field dataset ($n = 2001$) into two roughly equal halves—a *calibration* sample and a *holdout* sample. The first half, the calibration sample, was used in the initial model testing and modification phase, while the second-half, the holdout sample, was retained for later use when confirmation of the final measurement model was needed.

In order to split the dataset into calibration and holdout sub-samples, the final dataset was randomly stratified across the three different media. As a result of the stratification process and our restriction of equal sub-sample sizes, three observations were dropped from the dataset. This resulted in the calibration and holdout datasets each consisting of 999 observations (493 telephone users, 301 email users, and 205 instant messaging users apiece).

Using SAS proc CALIS (Hatcher, 1994), we assessed the fit of the original, hypothesized measurement model, using maximum likelihood (ML) estimation, against our observed data in the calibration sample. Proc CALIS produces several different measures of model "fit," all of which provide slightly different answers to the question, "how well does the covariance matrix estimated by the model match the covariance matrix observed in our empirical data?" These measures of fit generally fall into three categories.

First, the χ^2 *goodness-of-fit statistic* assesses "the magnitude of discrepancy between the sample and fitted covariance matrices, and it is the product of the sample size minus one and the minimum fitting function" (Hu and Bentler, 1999, p. 2). Unfortunately, the χ^2 statistic is sensitive to large sample sizes and will reject even a closely fitting model. Very small, even managerially inconsequential, differences may result in a rejection of the model based on the χ^2 statistic.

The second category consists of *absolute fit indices*, which "assesses how well an *a priori* model reproduces the sample data" (Hu and Bentler, 1999, p. 2). Proc CALIS produces estimates of several absolute fit indices, including the goodness of fit index (GFI) and the root mean squared error of approximation (RMSEA), which we have provided in the tables.

The third general category of fit index contains *incremental fit indices*, which measure "the proportionate improvement in fit by comparing a target model with a more restricted, nested baseline model"

Table 4
Results for CFA of initial measurement model—calibration sample ($n = 999$)

Fit statistic/index	Score	Recommended[a]
χ^2	1801.2 (389 df)[b]	Non-significant
Goodness of fit index (GFI)	0.89	0.90
RMSEA	0.06	0.06
RMSEA 90% CI	0.058–0.063	–
Bentler's CFI	0.96	0.95
Bentler and Bonett's NNI	0.95	0.95
Bollen's Normed Rho1	0.94	0.95
Bollen's Delta2	0.96	0.95

[a] See Browne and Cudeck, 1993; Jöreskog, 1993; Hu and Bentler, 1999.

[b] Significant at $P < 0.01$.

(Hu and Bentler, 1999, p. 2). Incremental fit indices produced by proc CALIS and included in our results tables include Bentler's Comparative Fit Index (CFI), Bentler and Bonett's Non-Normed Index (NNI), Bollen's Normed Rho1, and Bollen's Delta2. Generally, it is desirable to consider a combination of fit indices from all three of these categories when evaluating overall model fit, and we present several different measurements of fit for each model tested in our results tables.

The results of this first test are shown in Table 4. The fit measurements indicate that the model is generally adequate overall. However, some misfit is evident by virtue of the fact that many of our fit indices are right at or below the recommended values for "close fit." The significant χ^2 can be explained by virtue of the large sample size. The goodness of fit index (GFI) is similarly influenced by sample size (Gerbing and Anderson, 1993), with a tendency to move further away from 1.00 (ideal) as sample size increases. Thus, while it is still widely used, GFI is a less-than-ideal measure of global fit. The root mean square error of approximation (RMSEA) is not as susceptible to influence from our large sample size, and its value of 0.06 suggests acceptable fit.

Given these global results, we next examined the CFA results for evidence for reliability and convergent validity. All factor loadings (λ) were significant (at $P < 0.01$) and large (>0.50), indicating that the indicators did adequately reflect their corresponding constructs. Second, we examined scale coefficient ω scores for each scale (Edwards, 2003), and all scale ω scores surpassed the 0.70 threshold. Third, we examined the average variance extracted (AVE) for each scale (Fornell and Larcker, 1981), and all AVE scores surpassed the 0.50 threshold.

As satisfactory as these results appear (Appendix B), the global fit statistics above suggest that there may be opportunities for improving our measurement model slightly. We may want to explore the individual parameter estimates to determine which ones are significantly contributing to model misfit (Hu and Bentler, 1999). We did this by looking at two criteria.

First, in some instances, inclusion of an item resulted in an attenuation of the scale ω, suggesting that the item did not contribute to the internal reliability of the scale. So, if removing an item improved the scale ω (i.e. reliability would be increased), the item was noted for further scrutiny and future elimination from the scale.

Second, by examining the Lagrange multipliers (Chou and Bentler, 1990) for each estimated λ, we found that a handful of indicators exhibited varying degrees of multidimensionality (i.e. the model fit would improve significantly if these indicators were permitted to map onto more than one latent construct). This small group of multi-dimensional indicators contributed the most to overall measurement model misfit in the CFA. Appendix B lists these items under the scales upon which they were designed primarily to load, as well as several statistics relevant to the scale evaluation process. These multidimensional indicators were then noted for future removal from the measurement model.

Finally, prior to making any changes to the scale contents, we ran a series of pairwise discriminant validity tests on the *latent constructs* by fixing, one at a time, the covariances between latent variables to unity. After each covariance was so constrained, we then compared the resulting χ^2 statistic for the entire measurement model with that of the model where the covariance is freely estimated. Using the χ^2 difference test, we were able to measure the significance of the subsequent increase in model misfit (Bollen, 1989; Anderson, 1987). Every covariance, when fixed to one, resulted in a significant increase (at $P < 0.01$) in model misfit, suggesting that the latent constructs exhibit satisfactory discriminant validity. The Φ matrix of latent construct covariances is provided in Table 5.

Table 5
Covariances among latent constructs (Φ) for initial measurement Model

	IRB	LB	UB	DAB	IAB	AM	AE	AP	IM
LB	0.75								
UB	0.78	0.94							
DAB	0.54	0.53	0.62						
IAB	0.72	0.68	0.79	0.80					
AM	0.67	0.74	0.82	0.57	0.68				
AE	0.77	0.88	0.99	0.64	0.83	0.79			
AP	0.58	0.64	0.73	0.54	0.57	0.66	0.74		
IM	0.64	0.68	0.78	0.56	0.72	0.86	0.75	0.56	
IP	0.53	0.56	0.68	0.54	0.56	0.64	0.69	0.92	0.61

All covariances are significant at $P<0.01$.

This "cleansing" of the scales is an important and theoretically valuable step, as it helps ensure that the scales are as homogeneous, and the measures are as congeneric, as possible. It also helps ensure that our claims of both convergent and discriminant validity are reinforced. Since the indicators being eliminated were found *not* to be consistent with the other measures associated with any given construct, retaining them would reduce the expected level of discriminant validity in our scales (Pedhazur and Schmelkin, 1991). Furthermore, by eliminating these undesirable indicators, the remaining items for each scale should exhibit more shared variance as a set. Additionally, these changes to the measurement model did not violate the "Two Indicator Rule" of identification (Bollen, 1989), allowing us to proceed with testing the revised measurement model, again using CFA and the calibration dataset.

4.2. Revised measurement model analysis

The measurement model, now incorporating the modifications described in Section 4.1, was re-tested using the calibration sample with the results shown in Table 6. As can be seen, there is an improvement in all measures of global fit. Moreover, the measures of global fit indicate that the measurement model demonstrates adequate fit overall. We also recalculated all of the item-level and construct-level measures of reliability and validity detailed in Section 4.1. Coefficient ω and the AVE of each construct for this revised measurement model are shown in Appendix B, as are the various indicator reliability measures.

Additionally, we re-ran the pairwise discriminant validity tests. With one exception, all covariances, when set to unity, resulted in a significant increase in model misfit (again verified with the χ^2 difference test). The one exception was $\Phi_{\text{UB–AE}}$, the covariance between usefulness belief (UB) and attitude towards the contact episode (AE). Fixing that covariance to one resulted in a χ^2 of 493.3 (165 df), a non-significant increase (at $P < 0.01$) of 0.8 (1 df) over the unconstrained revised measurement model. Two interpretations can be made of this result.

The first interpretation is that these two constructs, usefulness belief (UB) and attitude towards the contact episode (AE), do not exhibit discriminant validity and are actually the same construct. The second interpretation is that there is a strong *causal* relationship between UB and AE, and this causal relationship creates very high covariances among the

Table 6
Global fit statistics for CFA of the revised measurement model using calibration sample ($n = 999$) and holdout sample ($n = 999$)

Fit statistic/index	Revised Model, calibration sample	Revised Model, holdout sample
χ^2	492.5 (164 df)[a]	656.6 (164 df)[a]
Goodness of fit index (GFI)	0.96	0.94
RMSEA	0.04	0.05
RMSEA 90% CI	0.040–0.049	0.051–0.059
Bentler's CFI	0.99	0.98
Bentler and Bonett's NNI	0.98	0.97
Bollen's Normed Rho1	0.97	0.96
Bollen's Delta2	0.99	0.98

[a] Significant at $P < 0.01$.

indicators for these two constructs. A measurement model like the one run here would not be able to discern between these two explanations, as either could account for the results we have observed. Thus, we believe that, in this situation, the causality-based explanation is the more plausible of the two for a few reasons.

The literature cited earlier leads one to expect a strong causal relationship between the UB and AE constructs, as shown in our general model (Fig. 2). There are also strong *conceptual* differences between these two constructs (i.e. they clearly tap into two different concepts), and these conceptual differences are carefully reflected in the wording of the questions employed for the observed indicators (see Appendix B). These theoretical and conceptual arguments seem to strongly support the causality-based explanation, whereas very little support is apparent for the "same construct" explanation. If the causality-based explanation is accepted, then discriminant validity is supported, as two constructs must be distinct from each other in order to exhibit a causal relationship.

4.3. Validation Using the Holdout Sample

Satisfied with the results of the revised measurement model using the calibration sample, we proceeded to re-run the CFA using the holdout sample. The results, shown in Table 6, showed similarly appropriate results. These results indicate that the final measurement model fits the holdout sample quite acceptably, indicating that the two samples exhibit *invariance of form* (i.e. it is appropriate to use the same mapping of observed variables onto latent variables for both sub-samples). This finding increases our confidence that the final measurement model is generalizable to Iotanet's entire customer base and that operational measures of the constructs defined in the hypothesized B–A–I conceptual framework depicted in Fig. 2 have sufficient psychometric properties for advancing research in technology-mediated contexts.

5. Conclusions

The primary contributions of this paper are the definition of new constructs associated with the technology-mediated customer service experience and the development of new multi-item measurement scales for measuring these constructs. Unlike much prior SOM research, our study takes a grounded theory approach using customers' perceptions. Future SOM empirical research linking these constructs in causal or structural models in a technology-mediated customer contact situation will benefit significantly from the existence of relevant construct definitions and good measurement scales. A secondary contribution of this work is the demonstration of a rigorous empirical scale and item development process.

Like any research, our approach and our results have some limitations. First, the use of convenience samples in the pre- and pilot-tests may have limited our insights early in the process. The use of random sampling in the final data analysis, however, alleviated much of the concern regarding this issue. A second limitation is the fact that our CFA resulted in some scales having only two indicators. While this may prove to be a limitation in some applications and some models, identification methods do exist that support their re-use in new models (Bollen, 1989). Moreover, the reader is provided with three or more items for each scale (with the exception of DAB) in Appendix B. While our CFA results indicated that our measurement model fared better with certain items dropped from our final analysis, future researchers may find that these specific items may perform adequately in their research contexts. A third potential limitation centers on the possibility of discriminant validity of the UB and AE constructs. While we feel that a reasonable explanation is that a strong causal relationship creates a large statistical covariance between these two constructs, future researchers who employ the UB and AE scales should pay particular attention to this potential issue when examining their results.

Finally, while CFA is generally the preferred method for establishing convergent and discriminant validity for a measurement model, it is not without weaknesses, two in particular are noted by Bagozzi et al. (1991). They indicate that the base structure of CFA has two potential problems. First, it confounds two different sources of variance (measurement error and unique true-score variance other than that explained by the traits and methods) in the observed measures. Second, it assumes that variation in indicator variables is strictly a linear combination of these sources of variance, thereby eliminating any opportunity for discovering trait-method interactions. Given the consistently high reliabilities of our scales, the first weakness should not be of considerable concern. Given that our scales are assessed using a single method, we have no opportunity for detecting whether the linearity constraint is a source of concern. This is a potential limitation of the study, but is consistent with other research that relies on CFA to support claims of construct validity.

Despite these issues, scale development, like research in general, is an iterative process. We feel our new multi-item measurement scales based on the B–A–I theoretic framework will provide SOM researchers with robust starting points when investigating technology-mediated customer contact and e-service operations from a customer's perspective, a relatively new line of research. Both academia and industry are likely to benefit from additional research on how customer contact can be most effectively managed in technology-mediated service environments, and these clearly defined constructs and robust measurement scales will significantly aid future researchers' investigations on this important topic.

Appendix A. Example of indicator testing and evolution

Using the information richness belief (IRB) construct as an example, Table A.1 shows how items for a scale were modified in each stage of the scale development process (see Fig. 3 for process flowchart).

Table A.1
Illustrative example

Construct and final item verbiage	Pre-test	Pilot #1	Pilot #2	Final survey
IRB—information richness belief				
How close to actual face-to-face interaction was your communication with Iotanet customer service?	X	X	X	X
How well did your communication with Iotanet customer service include the subtleties and nuances usually associated with face-to-face interaction?	X	X	X	–
How efficiently did your communication with the service provider convey the necessary information between you and the other person?	X	X	X	–
I believe it was easy to understand what the other person was talking about (or meant) during my communication with the service provider.	X	X	–	–
How much did your communication with the service provider feel like a face-to-face, in-person conversation?	–	–	X	X
How much did your communication with Iotanet customer service feel like a face-to-face, in-person conversation?	–	–	–	X

Appendix B. CFA results for measurement scales and associated indicators

Scales and associated indicators	Scale points	Upper anchor, lower anchor	Initial Model (CFA #1) $n = 999$			Revised Model (CFA #2) $n = 999$		
			Standardized path loading (λ)	Item reliability (R_m^2)	Item-to-total correlations	Standardized path loading (λ)	Item reliability (R_m^2)	Item-to-total correlations
IRB—Information richness belief			$\omega = 0.85$, average $= 0.66$			$\omega = 0.85$, average $= 0.66$		
How close to actual face-to-face interaction was your communication with Iotanet customer service?	7	Extremely, not at all	0.91	0.83	0.77	0.91	0.83	0.79
How much did your communication with Iotanet customer service feel like a face-to-face, in-person conversation?	7	Extremely, not at all	0.90	0.81	0.77	0.90	0.81	0.77
How much did your communication with the service provider feel like a face-to-face, in-person conversation?	7	Extremely, not at all	0.58	0.34	0.54	0.58	0.34	0.57
LB—learning belief			$\omega = 0.93$, average $= 0.76$			$\omega = 0.84$, average $= 0.72$		
After your communication with Iotanet customer service, how much more knowledgeable were you about your issue?	7	Extremely, not at all	0.91	0.82	0.84	0.89	0.79	0.71
I believe my communication with Iotanet customer service enabled me to look at the situation from a new perspective.	7	Strongly agree, strongly disagree	0.82	0.67	0.79	0.81	0.65	0.71
I believe my communication with Iotanet customer service was educational.	7	Strongly agree, strongly disagree	0.83	0.69	0.79	Dropped: multidimensionality		
I believe my communication with Iotanet customer service provided me with new knowledge.	7	Strongly agree, strongly disagree	0.93	0.87	0.89	Dropped: multidimensionality		
UB—usefulness belief			$\omega = 0.94$, average $= 0.80$			$\omega = 0.88$, average $= 0.78$		
I believe communicating with Iotanet customer service was a useful experience.	7	Strongly agree, strongly disagree	0.93	0.87	0.90	0.92	0.84	0.87
How much additional value did the experience of communicating with Iotanet customer service add for you?	7	Extremely, none	0.86	0.74	0.83	0.85	0.73	0.87
The experience of communicating with Iotanet customer service was how useful to you?	7	Extremely, none	0.91	0.83	0.86	Dropped: multidimensionality		
I believe that the experience of communicating with Iotanet customer service added value to the service.	7	Strongly agree, strongly disagree	0.88	0.77	0.84	Dropped: multidimensionality		

Appendix B (*Continued*)

Scales and associated indicators	Scale points	Upper anchor, lower anchor	Initial Model (CFA #1) $n = 999$			Revised Model (CFA #2) $n = 999$		
			Standardized path loading (λ)	Item reliability (R_m^2)	Item-to-total correlations	Standardized path loading (λ)	Item reliability (R_m^2)	Item-to-total correlations
DAB – Duration Appropriateness Belief			$\boldsymbol{\omega} = 0.80$, average $= 0.66$			$\boldsymbol{\omega} = 0.80$, average $= 0.66$		
I believe the time I spent actively communicating (talking to someone, reading or writing emails, chatting) with Iotanet customer service should have been:[a,b]	13	Much longer, much shorter	0.80	0.63	0.66	0.80	0.64	0.66
I believe the overall length of time I spent actively communicating (talking to someone, reading or writing emails, chatting) with Iotanet customer service was:[b]	13	Too long, too short	0.83	0.70	0.66	0.83	0.69	0.66
IAB – Intimacy Appropriateness Belief			$\omega = 0.78$, average $= 0.54$			$\omega = 0.79$, average $= 0.66$		
How impersonal did you consider your communication with Iotanet customer service to be?[a,b]	13	Too much, not enough	0.75	0.57	0.71	0.85	0.72	0.70
How intimate do you believe your communication with Iotanet customer service was?[b]	13	Too much, not enough	0.68	0.46	0.63	0.77	0.60	0.70
How friendly and personal was your communication with Iotanet customer service?	7	Extremely, not at all	0.78	0.62	0.56	Dropped: multidimensionality, attenuation of scale ω		
AM—attitude towards the medium			$\omega = 0.91$, average $= 0.78$			$\omega = 0.93$, average $= 0.87$		
I was pleased by using [this medium].	7	Strongly agree, strongly disagree	0.93	0.87	0.87	0.92	0.85	0.87
How satisfying was using [this medium]?	7	Extremely, not at all	0.94	0.88	0.85	0.95	0.90	0.87
I enjoyed using [this medium].	7	Strongly agree, strongly disagree	0.77	0.59	0.75	Dropped: multidimensionality, attenuation of scale ω		
AE—attitude towards the episode			$\omega = 0.93$, average $= 0.81$			$\omega = 0.94$, average $= 0.88$		
I was pleased by the experience of communicating with Iotanet customer service.	7	Strongly agree, strongly disagree	0.94	0.88	0.88	0.94	0.88	0.88
How satisfying was the experience of communicating with Iotanet customer service?	7	Extremely, not at all	0.94	0.88	0.87	0.94	0.88	0.88
How disappointing was the experience of communicating with Iotanet customer service?[a]	7	Extremely, not at all	0.81	0.66	0.79	Dropped: attenuation of scale ω		

Appendix B (*Continued*)

Scales and associated indicators	Scale points	Upper anchor, lower anchor	Initial Model (CFA #1) $n = 999$			Revised Model (CFA #2) $n = 999$		
			Standardized path loading (λ)	Item reliability (R_m^2)	Item-to-total correlations	Standardized path loading (λ)	Item reliability (R_m^2)	Item-to-total correlations
AP—attitude towards the service provider			$\omega = 0.94$, average $= 0.83$			$\omega = 0.95$, average $= 0.91$		
I am pleased with Iotanet	7	Strongly agree, strongly disagree	0.94	0.87	0.89	0.95	0.87	0.89
How satisfied are you with Iotanet?	7	Extremely, not at all	0.95	0.90	0.89	0.96	0.91	0.89
I enjoy using Iotanet.	7	Strongly agree, strongly disagree	0.84	0.71	0.81	Dropped: multidimensionality, attenuation of scale ω		
IM—Intention towards the medium			$\omega = 0.95$, average $= 0.86$			$\omega = 0.95$, average $= 0.86$		
I would use [this medium] again to contact Iotanet customer service.	7	Strongly agree, strongly disagree	0.92	0.85	0.88	0.92	0.85	0.88
I intend to use [this medium] the next time I need to contact Iotanet customer service.	7	Strongly agree, strongly disagree	0.94	0.89	0.91	0.94	0.89	0.90
How likely are you to use [this medium] again the next time you need to contact Iotanet customer service?	7	Extremely, not at all	0.92	0.84	0.88	0.92	0.84	0.87
IP—Intention Towards the Provider			$\omega = 0.94$, average $= 0.83$			$\omega = 0.92$, average $= 0.86$		
I intend to continue using Iotanet.	7	Strongly agree, strongly disagree	0.91	0.83	0.87	0.92	0.84	0.85
How likely are you to use Iotanet again?	7	Extremely, not at all	0.93	0.87	0.89	0.93	0.86	0.85
I intend to use Iotanet again in the future.	7	Strongly agree, strongly disagree	0.90	0.81	0.86	Dropped: multidimensionality		

All standardized path loadings are significant at $P < 0.01$.

[a] Item is reverse-coded.

[b] Item is bi-directional.

References

Adam Jr., E.E., Swamidass, P.M., 1989. Assessing operations management from a strategic perspective. Journal of Management 15 (2), 181–203.

Agarwal, R., Prasad, J., 1999. Are individual differences germane to the acceptance of new information technologies? Decision Sciences 30 (2), 361–391.

Alavi, M., 1994. Computer-mediated collaborative learning: an empirical evaluation. MIS Quarterly 18 (2), 159–174.

Alavi, M., Yoo, Y., Vogel, D.R., 1997. Using information technology to add value to management education. Academy of Management Journal 40 (6), 1310–1333.

Anderson, J.C., 1987. An approach for confirmatory measurement and structural equation modeling of organizational properties. Management Science 33 (4), 525–542.

Anderson, J.C., Gerbing, D.W., 1991. Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. Journal of Applied Psychology 76 (5), 732–740.

Ajzen, I., 1985. From Intentions to Actions: A Theory of Planned Behavior. In: Kuhl, J., Beckman, J. (Eds.), Action-Control: From Cognition to Behavior, Heidelberg, Springer, pp. 11–39.

Ajzen, I., 1991. The theory of planned behavior. Organizational Behavior and Human Decision Processes 50, 179–211.

Bagozzi, R.P., Yi, Y., Phillips, L.W., 1991. Assessing construct validity in organizational research. Administrative Science Quarterly 36, 421–458.

Barnes, J.G., Dunne, P.A., Glynn, W.J., 2000. Self-Service and Technology: Unanticipated and Unintended Effects on Customer Relationships. In: Swartz, Teresa A. Iacobucci, D. (Eds.), Handbook of Services Marketing and Management, Sage Publications, Thousand Oaks, CA, pp. 89–102.

Boland Jr, R.J., Tenkasi, R.V., 1995. Perspective making and perspective taking in communities of knowing. Organization Science 6 (4), 350–372.

Bollen, K.A., 1989. Structural Equations with Latent Variables. Wiley, New York, NY.

Boyer, K.K., Hallowell, R., Roth, A.V., 2002. E-Services: operating strategy—a case study and a method for analyzing operational benefits. Journal of Operations Management 20 (2), 175–188.

Browne, M.W., Cudeck, R., 1993. Alternative Ways of Assessing Model Fit. In: Bollen, K.A., Long, J.S. (Eds.), Testing Structural Equation Models. Sage, Newbury Park, CA, pp. 136–162.

Carlson, J.R., Zmud, R.W., 1999. Channel expansion theory and the experiential nature of media richness perceptions. Academy of Management Journal 42 (2), 153–170.

Carmines, E.G., Zeller, R.A., 1979. Reliability and Validity Assessment. Sage, Newbury Park, CA.

Chase, R.B., 1978. Where does the customer fit in the service operation? Harvard Business Review 56, 137–142.

Chase, R.B., 1981. The Customer Contact Approach to Services: Theoretical Bases and Practical Extensions. Operations Research 29 (4), 698–706.

Chou, C.-P., Bentler, P.M., 1990. Model modification in covariance structure modeling: a comparison among likelihood ratio, Lagrange multiplier, and Wald tests. Multivariate Behavioral Research 25 (1), 115–136.

Churchill, Gilbert A., 1995. Marketing Research: Methodological Foundations, Dryden Press, Forth Worth TX.

Coffee, P., 2001. E-CRM: making the customer king? eWeek 18 (20), 39–44.

Collier, D.A., 1983. The service sector revolution: the automation of services. Long Range Planning 16 (6), 10–20.

CommerceNet On-Line Research Report, 2000. Worldwide Internet Population, Internet Location: http://www.commerce.net/research/stats/wwstats.html.

Daft, R.L., Lengel, R.H., 1984. Information richness: a new approach to managerial information processing and organizational design. In: Staw, B.M., Cummings, L. (Eds.), Research in Organizational Behavior, JAI Press, Greenwich, CT, pp. 191–233.

Daft, R.L., Lengel, R.H., 1986. Organizational information requirements, media richness, and structural design. Management Science 32 (5), 554–571.

Davis, F.D., Bagozzi, R.P., Warshaw, P.R., 1989. User acceptance of computer technology: a comparison of two theoretical Models. Management Science 35 (8), 982–1003.

DeSanctis, G., Monge, P., 1999. Communication processes for virtual organizations. Organization Science 10 (6), 693–703.

Edwards, J.R., 2001. Multidimensional constructs in organizational behavior research: an integrative analytical framework. Organizational Research Methods 4, 144–192.

Edwards, Jeffrey R., 2003. Construct Validation in Organizational Behavior Research. In: Greenberg, J. (Ed.), Organizational Behavior: The State of the Science, 2nd edition. Erlbaum, Mahwah, NJ.

Fishbein, M., Azjen, I., 1975. Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research. Addison-Wesley, Reading, MA.

Fitzsimmons, J.A., Fitzsimmons. M.J., 2004. Service Management: Operations, Strategy, and Information Technology. Irwin-McGraw Hill, New York, NY.

Flynn, B.B., Sakakibara, S., Schroeder, R.G., Bates, K.A., Flynn, J.E., 1990. Empirical research method in operations management. Journal of Operations Management 9 (2), 250–284.

Fornell, C., Larcker, D.F., 1981. Evaluating Structural Equation Models with unobservable variables and measurement error. Journal of Marketing Research 18 (1), 39–50.

Fowler, Jr., F.J., 1988. Survey Research Methods. Sage Publications, Newbury Park, CA.

Frohlich, M.T., Westbrook, R., 2002. Demand chain management in manufacturing and services: web-based integration, drivers and performance. Journal of Operations Management 20, 729–745.

Gerbing, D.W., Anderson, J.C., 1993. Monte Carlo Evaluations of Goodness-of-Fit Indices for Structural Equation Models. In: Bollen, K.A., Long, J.S. (Eds.), Testing Structural Equation Models. Sage Publications, Newbury Park, CA, pp. 40–65.

Ghutek, B.A., Bhappu, A.D., Liao-Troth, M., Cherry, B., 1999. Distinguishing between service relationships and encounters. Journal of Applied Psychology 84 (2), 218–233.

Hatcher, L., 1994. A Step-by-Step Approach to Using the SAS System for Factor Analysis and Structural Equation Modeling. SAS Institute, Inc., Cary, NC.

Haynes, R.M., Thies, E.A., 1991. Management of technology in service firms. Journal of Operations Management 10 (3), 388–397.

Heim, G.R., Sinha, K.K., 2001a. Operational drivers of customer loyalty in electronic retailing: an empirical analysis of electronic food retailers. Manufacturing and Service Operations Management 3 (3), 264–271.

Heim, G.R., Sinha, K.K., 2001b. A product-process matrix for electronic B2C operations: implications for the delivery of customer value. Journal of Service Research 3 (4), 286–299.

Hensley, R.L., 1999. A review of operations management studies using scale development techniques. Journal of Operations Management 17, 343–358.

Heskett, J.L., Jones, T.O., Loveman, G.W., Sasser, W.E., Schlesinger, L.A., 1994. Putting the service-profit chain to work. Harvard Business Review 72 (2), 164–174.

Heskett, J.L., Sasser, W.E., Schlesinger, L.A., 2003. The Value Profit Chain, The Free Press, NY.

Hill, A.V., Collier, D.A., Froehle, C.M., Goodale, J.C., Metters, R.D., Verma, R., 2002. Research opportunities in service process design. Journal of Operations Management 20 (2), 189–202.

Hu, L.-t., Bentler, P.M., 1999. Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. Structural Equation Modeling 6 (1), 1–55.

Huete, L., Roth, A.V., 1988. The industrialization and span of retail banks' delivery systems. International Journal of Operations and Production Management 8 (3), 46–66.

Jöreskog, Karl G., 1993. Testing Structural Equation Models. In: Bollen, K.A., Long, J.S. (Eds.), Testing Structural Equation Models, Sage, Newbury Park, CA, pp. 294–316.

Kellogg, D.L., Chase, R.B., 1995. Constructing an empirically derived measure for customer contact. Management Science 41 (11), 1734–1749.

Miceli, M.P., Jung, I., Near, J.P., Greenberger, D.B., 1991. Predictors and outcomes of reactions to pay-for-performance plans. Journal of Applied Psychology 76, 508–521.

Menor, L.J., 2000, An Empirical Investigation of New Service Development Competence and Performance. Unpublished Doctoral Dissertation, University of North Carolina, Chapel Hill.

Nie, N.H., Erbing, L., 2000. Internet and Society: A Preliminary Report. Stanford Institute for the Quantitative Study of Society (SIQSS), Stanford, CA.

Nunnaly, J.C., 1978. Psychometric Theory. McGraw-Hill, New York, NY.

Oliveira, P., Roth, A.V., Gilland, W., 2002. Achieving competitive capabilities in E-services. Technological Forecasting and Social Change 69 (7), 721–731.

Parasuraman, A., Zeithaml, V.A., Berry, L.A., 1988. SERVQUAL: a multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing 64 (1), 12–40.

Parzinger, Monica J., 1999. Factors Influencing Response Rates in Web-Based Surveys. In: Proceedings of the 1999 Annual Meeting of the Decision Sciences Institute, pp. 686–688.

Pedhazur, E.J., Schmelkin, L.P., 1991. Measurement, Design, and Analysis: An Integrated Approach, Earlbaum, Hillsdale, NJ.

Pine, B.J., Gilmore, J.H., 1998. Welcome to the experience economy. Harvard Business Review 76 (4), 97–105.

Podsakoff, P.M., Organ, D.W., 1986. Self reports in organizational research: problems and prospects. Journal of Management 12, 531–544.

Pontes, M.C.F., Kelly, C.O., 2000. The identification of inbound call center agents' competencies that are related to callers' repurchase intentions. Journal of Interactive Marketing 14 (3), 41–49.

Porter, M.E., Millar, V.E., 1985. How information gives you competitive advantage. Harvard Business Review 63 (4), 149–160.

Reichheld, F.F., Schefter, P., 2000. E-loyalty: your secret weapon on the web. Harvard Business Review 78 (4), 105–113.

Roth, A.V., 2000. Service Strategy and the Technological Revolution: The 7 Myths of E-Services. In: Machuca, J.A.D., Mandakovic, T. (Eds.), POM Facing the New Millennium: Evaluating the Past, Leading with the Present and Planning the Future of Operations. POM Sevilla, pp. 159–168.

Roth, A.V., 2001. Siete Mitos De Los Servicios electronic (Seven Myths of Electronic Services). Harvard Deusto Business Review 100, 88–96.

Roth, A.V., Menor, L.J., 2004. Insights into services operations management: a research agenda. Productions and Operations Management 12 (2), 145–163.

Roth, Aleda V., Schroeder, Roger. Scales in Operations Management. Sage, Thousand Oaks, CA, in press.

Sasser, W.E., Olsen, R.P., Wyckoff, D.D., 1978. Management of Service Operations: Text, Cases, and Readings. Allyn and Bacon, Boston, MA.

Soteriou, A.C., Chase, R.B., 1998. Linking the customer contact model to service quality. Journal of Operations Management 16, 495–508.

Stratman, J.K., Roth, A.V., 2002. Enterprise resource planning (ERP) competence constructs: two-stage multi-item scale development and validation. Decision Sciences 33 (4), 601–628.

Walley, P., Amin, V., 1994. Automation in a customer contact environment. International Journal of Operations and Production Management 14 (5), 86–100.

Wiersema, Fred, 1996. Customer Intimacy. Knowledge Exchange. Santa Monica, CA.

Zeithaml, V.A., Parasuraman, A., Malhotra, A., 2002. Service quality delivery through web sites: a critical review of extant knowledge. Journal of the Academy of Marketing Science 30 (4), 362–375.