University of Wollongong
Research Online

Faculty of Commerce - Papers (Archive) Faculty of Business and Law

2007

Toward a valid Measure of e-retailing service quality

John R. Rossiter
University of Wollongong, jrossite@uow.edu.au

Follow this and additional works at: https://ro.uow.edu.au/commpapers

 Part of the Business Commons, and the Social and Behavioral Sciences Commons

Recommended Citation

Rossiter, John R.: Toward a valid Measure of e-retailing service quality 2007, 36-48.
https://ro.uow.edu.au/commpapers/1269

Research Online is the open access institutional repository for the University of Wollongong. For further information contact the UOW Library: research-pubs@uow.edu.au

Toward a valid Measure of e-retailing service quality

Abstract

E-retailers are major players in the field of electronic commerce and their success would seem to depend on service quality, because they are selling the same products that traditional retailers sell. This article critiques Collier and Bienstocks [5] new measure of e-retailing service quality and shows how the stages of e-retailing service quality can be more validly measured by adopting Rossiters [12] C-OAR-SE procedure for scale development. Collier and Bienstock`s measure is insufficiently valid because the measure (1) fails to specify the hierarchical objects that form the construct, and measures the overall object, e-retailing, wrongly by focusing on completed transactions; (2) does not fully acknowledge the hierarchy of attributes that form the construct and operationalizes these attributes wrongly as reflective when at all four levels they are formed; (3) inappropriately represents the rater entity by using college student participants; (4) employs unnecessarily numerous, often redundant, and sometimes ambiguous scale items, with Likert-type answer scales that make the observed scores managerially almost uninterpretable; and (5) tries to measure overall e-retailing service quality when it makes sense only to measure the separate quality ratings of sequential stages of the e-retailing service process. The article points out how these problems could be avoided by constructing a new measure that properly applies the C-OAR-SE procedure.

Keywords

Toward, valid, Measure, retailing, service, quality

Disciplines

Business | Social and Behavioral Sciences

Publication Details

Rossiter, J. R. (2007). Toward a valid Measure of e-retailing service quality. Journal of Theoretical and Applied Electronic Commerce Research, 2 (3), 36-48.

This journal article is available at Research Online: https://ro.uow.edu.au/commpapers/1269

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 2 / ISSUE 3 / DECEMBER 2007 / 36 - 48
© 2007 Universidad de Talca - Chile

This paper is Available online at
www.jtaer.com

Toward a Valid Measure of E-Retailing Service Quality

John R. Rossiter

[1] University of Wollongong, Marketing Research Innovation Centre, john_rossiter@uow.edu.au

Received 13 March 2007; received in revised form 15 May 2007; accepted 30 June 2007

Abstract

E-retailers are major players in the field of electronic commerce and their success would seem to depend on service quality, because they are selling the same products that traditional retailers sell. This article critiques Collier and Bienstock's [5] new measure of e-retailing service quality and shows how the stages of e-retailing service quality can be more validly measured by adopting Rossiter's [12] C-OAR-SE procedure for scale development. Collier and Bienstock`s measure is insufficiently valid because the measure (1) fails to specify the hierarchical objects that form the construct, and measures the overall object, e-retailing, wrongly by focusing on completed transactions; (2) does not fully acknowledge the hierarchy of attributes that form the construct and operationalizes these attributes wrongly as "reflective" when at all four levels they are "formed"; (3) inappropriately represents the rater entity by using college student participants; (4) employs unnecessarily numerous, often redundant, and sometimes ambiguous scale items, with Likert-type answer scales that make the observed scores managerially almost uninterpretable; and (5) tries to measure overall e-retailing service quality when it makes sense only to measure the separate quality ratings of sequential stages of the e-retailing service process. The article points out how these problems could be avoided by constructing a new measure that properly applies the C-OAR-SE procedure.

Key words: C-OAR-SE procedure, e-retailing service quality, scale development

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 2 / ISSUE 3 / DECEMBER 2007 / 36 - 48
© 2007 Universidad de Talca - Chile

This paper is Available online at
www.jtaer.com

1 Introduction

E-retailing is the largest marketing activity in the rapidly growing field of electronic commerce and, logically, perceived service quality would seem to be the key success factor that lifts this new form of retailing above traditional retailing because the products the two types of retailers sell are the same. However, e-retailing service quality must be validly measured if its actual role is to be demonstrated empirically. Moreover, a valid measure of service quality at each stage of e-retailing is required if e-retailers seek to improve their service quality.

It is therefore crucial that a valid instrument for measuring e-retailing service quality in its main stages be developed. The new instrument needs to be very highly content valid [12] because only then can it be used to test theory, such as whether e-retailing service quality really does cause repeat patronage, and only then can it be useful in practice, assuming that empirical research confirms the theory that it does cause repeat patronage, by showing e-retail managers precisely what components of service quality need to be changed, and by how much, to retain customers [14], [18] As well, the new scale, which will actually be several scales, needs to be as concise as possible without sacrificing coverage of the main components, because only a relatively concise scale will be used by practitioners and by those academics who can't afford the questionnaire space that a lengthy instrument requires.

1.1 Collier and Bienstock's New Measure

Collier and Bienstock [5] developed a new multiple-item measure of e-retailing service quality which they claim to be more valid than previous measures. Their instrument consists of 54 items. They claim that it has better content validity than the academic E-S-QUAL and E-RecS-QUAL scales developed by Parasuraman, Zeithaml, and Malhotra [10] and the practitioner scales used by several leading internet rating services such as Consumer Reports' E-Ratings, BizRate.com, and Worldbestwebsites.com. Collier and Bienstock's new instrument consists of three separate scales, compared with Parasuraman, Zeithaml, and Malhotra's two, measuring three hypothesized stages of e-retailing service. Despite referring to e-retailing service quality overall, they do not use the full scale, the 54 items, to compute a total e-retailing service quality score. The separate scales refer to sequential stages of e-retailing, which they call *process* (which really refers only to the website visit), *outcome* (receipt of the delivered product), and *recovery* (a stage which occurs only if a complaint is made to the e-retailer).

1.2 C-OAR-SE Critique

The present article applies the C-OAR-SE procedure for scale development [12] to demonstrate that Collier and Bienstock's instrument for measuring e-retailing service quality [5] is not sufficiently content-valid and explains how a highly valid and briefer instrument can be developed. The C-OAR-SE procedure is also used to propose a more valid measure of Collier and Bienstock's ultimate dependent variable, future patronage intention. Rather ironically in light of the present critique, Collier and Bienstock attempted to go beyond the conventional scale development procedure [4] applied in most previous academic measures of service quality, notably the SERQUAL types of measures, by applying the C-OAR-SE procedure. However, as the present article explains, Collier and Bienstock in fact followed the conventional procedure and made the same mistakes as did the previous academic researchers.

C-OAR-SE (an acronym for Construct definition, Object classification, Attribute classification, Rater identification, Scale formation, and Enumeration) is a rational, expert judgment procedure, not an empirical, statistical one. It can be shown logically, using C-OAR-SE, independently of the empirical data they report, that Collier and Bienstock's measure is not sufficiently valid.

To develop a highly valid measure of e-retailing service quality, the full C-OAR-SE procedure should have been applied, in six steps, which form the headings of the present article:

1. Definition and measurement of the object of the construct.
2. Definition and measurement of the attribute of the construct.
3. Definition and sampling of the rater entity.
4. Content coverage of the formative items for the component attributes.
5. Content of the question part and the answer part of the items.
6. Enumeration (scoring) to improve predictive validity and managerial usefulness.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 2 / ISSUE 3 / DECEMBER 2007 / 36 - 48
© 2007 Universidad de Talca - Chile

This paper is Available online at
www.jtaer.com

2 Definition and Measurement of the Object of the Construct

According to C-OAR-SE, a construct comprises three essential elements: "(1) the object, including its constituents or components, (2) the attribute, including its components, and (3) the rater entity" (see [12], p. 308).

2.1 Overall Object

The overall object in this construct is *e-retailing*, which may seem obvious, but what is less obvious is that *e-retailing* is a fourth-order, abstract, collective, formed object. (Most readers of the C-OAR-SE article accept the idea of an abstract formed attribute because this dates back to the seminal distinction by Blalock [2] and Bollen and Lennox [3] and has been well publicized in the social sciences literature. But most haven't realized that objects, too, can be formed.) The object, *e-retailing*, is abstract because the term "e-retailing" could suggest different types of e-retailers to different raters as well as different stages of the overall e-retailing process (such as any of the stages identified by Collier and Bienstock [5]). The object *e-retailing* is collective of the types of e-retailers as constituents. The object *e-retailing* is formed because the stages are component objects that add up to the total object, although theoretically and practically it does not make much sense to perform this addition.

2.2 Stage Components

Collier and Bienstock [5], in their expansion of the construct, identify the likely stage components but they mistakenly describe these as components of the attribute, *service quality*, instead of components of the object, *e-retailing*. The stage components of *e-retailing*, renamed more accurately than Collier and Bienstock named them, are (1) the *website* visit; (2) the *transaction* with the e-retailer, if the visit culminates in a transaction or an attempted transaction; and (3) *recovery*, if a transaction has ensued and if the customer complains about a problem with delivery or with the merchandise or service purchased. Collier and Bienstock correctly argue that Parasuraman, Zeithaml, and Malhotra [10] unjustifiably collapse the first two stages into one, in E-S-QUAL. The present author contends that there are, in e-retailing, five potential stages: (1) the *website* visit; (2) the *transaction attempt* (which too often is aborted, as found by Sismeiro and Bucklin [15]); (3) the *customer assistance phone contact* (the usual response when an online transaction attempt fails); (4) the *transaction outcome* (which is the actual stage that Collier and Bienstock measured); and (5) *recovery* (if needed). However, the present critique will proceed as though Collier and Bienstock's three stages are adequate.

2.3 Constituents of the Object

Collier and Bienstock [5] sampled (and thus measured) the constituents of the object wrongly. They sampled as the constituents of the *e-retailing* object "the last e-retailer with which [each customer in the survey] enacted a transaction" (p. 267, emphasis added). Restriction of the sample to (completed) transactions must underrepresent the type of e-retailer whose site is primarily accessed for pre-purchase "information" and then the transaction is enacted offline via another direct response medium, such as going to a travel agent or telephoning an airline direct to purchase airline tickets after checking schedules and prices on airlines' websites. Another constituent type of e-retailer underrepresented would be those companies whose websites are primarily accessed for "entertainment" even though you can purchase their products online. Also, providers of search engines and recommender systems are often an important constituent of e-retailing (Steckel, Winer, Bucklin, Dellaert, Dreze, Haubl, Jap, Little, Meyvis, Montgomery, and Rangaswamy [16]) and search engines are mentioned as a constituent by Collier and Bienstock in their conceptual discussion of e-retailing (p. 264) but their instrument has no item to represent them. The conceptual definition of the object *e-retailing* requires that the constituent types of e-retailers be specified and sampled.

Less than 2% of e-retail website visits result in a purchase [19]. Collier and Bienstock's sampling of completed transactions means the probable omission of e-retailers whose websites were of such poor "quality" that the visitor did not proceed to the transaction stage. This omission would have the effect of truncating (cutting off the bottom of) the distribution of *website service quality* (Collier and Bienstock's *process quality*) scores in their empirical results and making it less likely that website service quality would predict any further behavior such as whether, in fact, the *transaction* (Collier and Bienstock's *outcome*) stage occurs.

2.4 The Transaction Stage

Collier and Bienstock [5] also measured one of the component objects of the overall object wrongly. In their operational measure of the *transaction* stage object, some of the questions (items 27, 28, 32 and 33 in particular, in the Appendix to Collier and Bienstock's article and reproduced in the present Appendix, later) do not allow this to be the first transaction with the e-retailer. These questions instead assume that the consumer is already a repeat customer and so not all of the items for the *transaction* stage can be validly answered by a first-time customer. It may be noted that many of the *fulfillment* items in Parasuraman, Zeithaml, and Malhotra's [10] measure of E-S-QUAL have the same problem.

2.5 Correct Procedure for Object Measurement

From the perspective of the C-OAR-SE procedure, the fourth-order object, *e-retailing*, should have been sampled from the experiences of customers, first-time and repeat, of a representative collection of constituent types of e-retailers. The component objects, which are third-order and remain the same at the second-order level of objects, namely the *website* visit, the *transaction*, and *recovery*, should have been explicitly recognized as separate objects, rather than as part of separate attribute descriptions, so that experts could more readily decide whether these objects constitute the correct stages of e-retailing. In fact, these separate objects share the same attribute which is, of course, *service quality*. The nature and measurement of the attribute of the construct is considered next.

3 Definition and Measurement of the Attribute of the Construct

3.1 Overall Attribute

The overall attribute of the construct *e-retailing service quality as perceived by college student customers* (who are the rater entity, see section 3) is *service quality*. As realized by Collier and Bienstock [5] when they cited the C-OAR-SE article, perhaps helped by the article's discussion of how to classify *service quality* as an attribute, *service quality* is a higher-order, formed attribute. However, Collier and Bienstock consider *service quality* to be second-order formed whereas, just as was the case for *e-retailing* as an object, it is actually a fourth-order formed attribute. It is formed (made up of, potentially) by summing the scores from several third-order components. These third-order components differ according to the component object being rated, not the attribute, *service quality* which is the same for all three, and are designated in the present article as *website service quality*, *transaction service quality*, and *recovery service quality*.

Where the "formed" idea comes in, according to Collier and Bienstock, is that the attribute scores on these three third-order components could be summed to form the total *service quality* attribute score. However, the three component scores do not actually need to be summed. It only makes sense to do so in the case of customers who experience the next two stages of e-retailing after a website visit. Even then, the separate scores resulting from treating the stage quality measures as separate scales would be more diagnostic than a summed score. Collier and Bienstock do not sum them in their empirical study and therefore never actually apply *service quality* as a formed attribute despite calling it one.

3.2 Misapplication of Formed Attribute Theory

The conceptual error that Collier and Bienstock [5] make thereafter is to claim to follow formed-attribute theory as in the C-OAR-SE procedure, but to depart from it when operationalizing the measure. Perhaps without realizing it, they ended up following conventional theory, no differently from the previous researchers in service quality. They made two mistakes. The initial mistake was in the way they measured the 11 second-order component attributes. The second-order component attributes, renamed more accurately here with their objects, are *website ease of use*, *website privacy, website design quality, website information trustworthiness, website functionality* (all relating to the *website* object); *condition of received order, timeliness of received order, accuracy of received order* (all relating to the *transaction* object); and *recovery personnel competence, recovery procedure quality*, and *recovery outcome fairness* (all relating to the *recovery* object). Collier and Bienstock identified the 11 components a priori but then they took all of the items in their initial battery of items and threw them into a factor analysis, with orthogonal factor rotation. It is simply not believable that the same 11 hypothesized components would emerge as 11 orthogonal factors. If this is what they did, and their description of the statistical procedure (p. 268) is too vague to tell, the temptation to reidentify the components ex post from the factors must have been great. Alternatively, what they may have done is looked for three factors corresponding to the third-order stage quality components. This would be just as incorrect conceptually, for two reasons. One is that there is no reason why the stage quality factors should be uncorrelated (orthogonal) or even that they should be factors if formed-attribute theory (a component theory not a factor theory) is followed. The other reason is that the five, three, and three second-order components that make up (form) the three respective third-order stage components do not have to unidimensionally "load" on them. Factor analysis is a statistical procedure that ignores the conceptual requirement that the five, three, and three components form the respective stage quality attribute scores and the components of a formed attribute do not have to be, and indeed should not be, unidimensional (see [12], p. 315). Thus, Collier and Bienstock intended to follow formed-attribute theory, as in C-OAR-SE, but in truth applied the conventional eliciting-attribute theory in deriving the 11 components, just like the previous researchers did for their components.

The second mistake made by Collier and Bienstock, and by the previous researchers, is to regard the 11 second-order component attributes themselves as eliciting (warranting "reflective" indicators in conventional terms, see p. 267 of their article) and failing to realize that these, too, are formed attributes. For example, the second-order component object and attribute *website ease of use* is clearly made up of the ratings on its items and does not cause them as the reflective indicator specification adopted by Collier and Bienstock assumes. The items, in turn, in some cases redundantly, represent first-order components, each with its own object and attribute. In wrongly assuming that the 11 second-order component attributes were eliciting ("reflective"), Collier and Bienstock committed the

conventional mistake (see [12], p. 315) of computing coefficient alpha and deleting potentially defining items. It may be noted that this same mistake was made in E-S-QUAL and E-RecS-QUAL earlier [10] and in this regard Collier and Bienstock's measure is no better. The "formative" versus "reflective" distinction is not just a minor difference of statistical treatment but rather a major conceptual choice that, if made wrongly, reduces the content validity of the entire measure.

3.3 Correct Procedure for Attribute Measurement

What should have been conceptualized and measured, according to C-OAR-SE theory, is a totally formed hierarchy: the first-order item scores form the second-order attribute component scores; the second-order scores form the third-order attribute component scores; and those scores, in turn, form the first-order overall *service quality* attribute score (if it is appropriate to use the overall sum score). The complex structure of the attribute and its components, as well as their object components, as they should have been conceptualized, is shown in Figure 1.

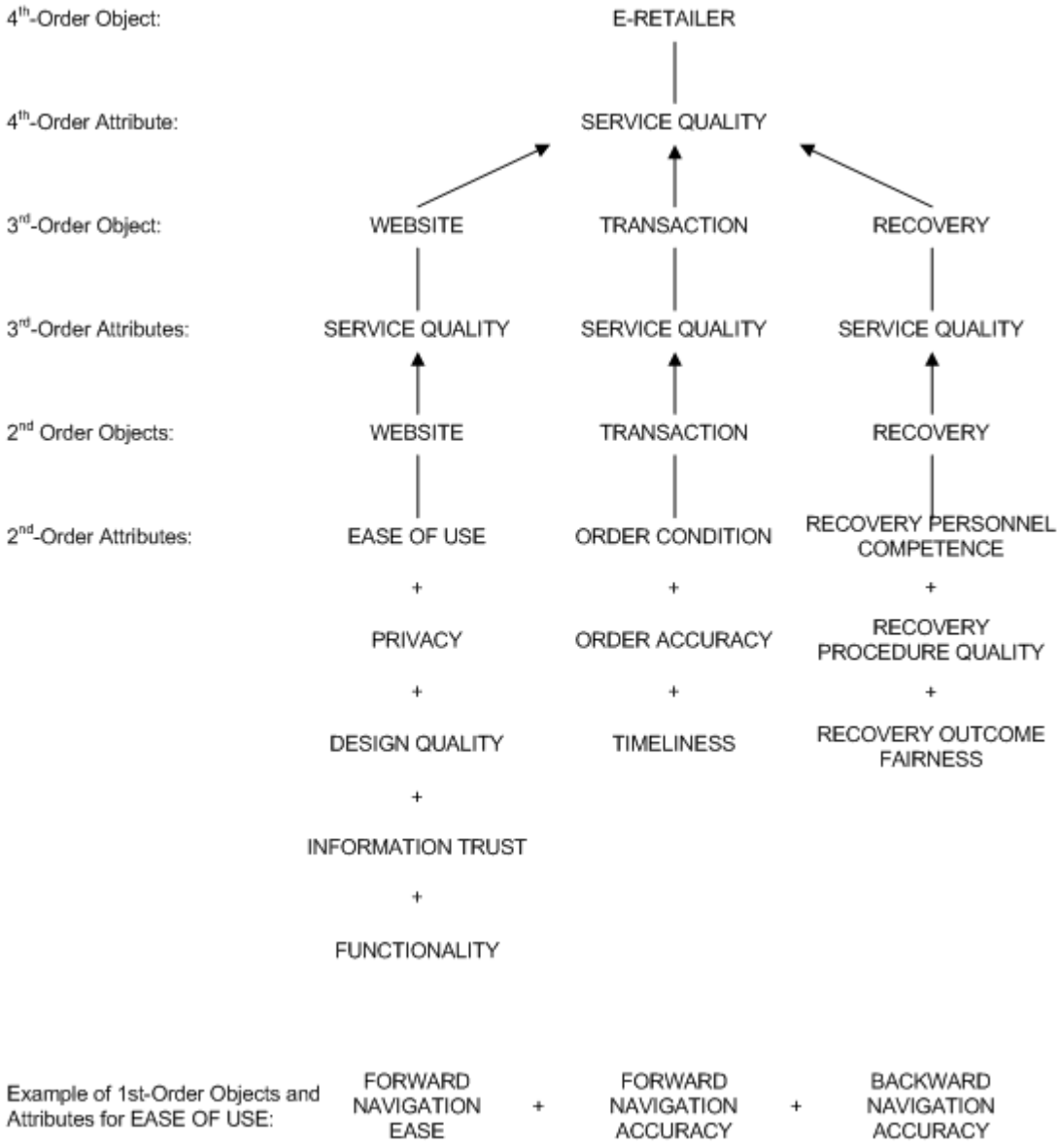


Figure 1: Construct Structure of *E-Retailing Service Quality* (Rater Entity: Unspecified) with First-Order Example

4 Definition and Sampling of the Rater Entity

4.1 Rater Entity

The third essential element in the conceptual definition of a construct, according to C-OAR-SE, is the rater entity. Critics of the C-OAR-SE procedure have argued against the inclusion of the rater entity in the construct (see especially [6]) but the Collier and Bienstock study [5] shows why it is essential. The complete label for the construct that Collier and Bienstock measured is *e-retailing service quality as perceived by college student customers*. The rater entity, *college student customers*, makes it a different construct than if the rater entity were, for example, the e-retail *manager* or an *expert panel*. If they were the same construct, then, logically, the measures of them would have to produce identical scores, regardless of the rater entity. This could happen empirically in a given dataset if all three rater entities happened to have the same perceptions, but there is no theoretical reason why they should. From the perspective of C-OAR-SE, consumers' quality ratings are one construct and managers' quality ratings and experts' quality ratings are others.

4.2 Wrong Sample of Raters

Collier and Bienstock [5] made an operational mistake when sampling the rater entity because they clearly intended the rater entity, the target population of raters, to be e-retail *customers* in general (their abstract and discussion imply this) but they settled for *young adult* e-retail customers and did not sample this narrower population properly. They chose a college student sample of raters because "the young adult population was the most active Web users [and] that sampling college students will allow us the best chance to represent the characteristics of online consumers" (p. 267). This choice is questionable because college students do not represent a random sample of young adults or even of frequent site-user young adults. In any case, Collier and Bienstock's justification for sampling young adults was that they are the most active web users, not necessarily the most active users of e-retailers. From the rater-entity perspective of C-OAR-SE, Collier and Bienstock measured a different construct, namely *e-retailing service quality as perceived by college student customers*. Importantly, Collier and Bienstock's college student sample is not a broad enough sample of raters upon which to test theoretical relationships or make empirical generalizations about e-retailing service quality.

4.3 Correct Sampling of the Rater Entity

A random sample of all e-retail customers should be taken for the rater entity and, if the researchers wish to learn about e-retailing success factors, the sample should be stratified by non-customers (website visitors only), first-time customers, occasional customers, and frequent customers of the e-retailer.

5 Content Coverage of the Formative Items for the Component Attributes

5.1 Right Stages?

Collier and Bienstock [5] identified the third-order objects and attributes of *website service quality*, *transaction service quality*, and *recovery service quality*, recalling that these are renamed in the present article, "from the analysis of both academic and practitioner literature" (p. 263). One might question their apparent need to rely on the literature to identify the object components of the formed object *e-retailing* when common knowledge would have sufficed. Collier and Bienstock are undoubtedly themselves customers of e-retailers and introspection would have suggested that contact with e-retailers has at least three separate stages, the last two of which are only potential: interaction on the website, which may or may not culminate in a purchase transaction; the transaction itself, if there is a transaction made online; and, if a problem occurs with the outcome, and the customer reports it, there will be a recovery stage. Where Collier and Bienstock made a contribution was in distinguishing the three stages (although the present author believes they should have distinguished five – see section 2.2 earlier). Parasuraman, Zeithaml, and Malhotra's E-S-QUAL measure [10], for example, wrongly included the *transaction* stage in the *website* visit stage, as the *fulfillment* factor, when clearly there may not be any fulfillment.

5.2 Right Second-Order Components?

A further problem arises with Collier and Bienstock's [5] identification of the 11 second-order attribute components, listed earlier and in Figure 1 earlier, five of which are nested under the *website service quality* component, three under the *transaction service quality* component, and three under the *recovery service quality* component. These 11 "dimensions" (and there is a problem right here because the word "dimension" implies that the attribute component is "reflective" when it is in fact formed and should be called a "component") are again claimed to be "based on practitioner and academic literature" (p. 264). However, in their summary of the previous literature in their Table 1 (p. 263), they merely list, separately, the components that previous academic and practitioner researchers have used or

currently use and make no apparent attempt to synthesize the lists and to derive a comprehensive set of mutually exclusive components. Six of the seven practitioners' component lists refer only to website quality, not to overall e-retailing service quality, as does one of the two academic research teams' component lists. Moreover, there appear to be several important component attributes of the *website* object in their Table 1 that do not appear in their final list; for example, system availability of the website (one of the components in E-S-QUAL), or range of product (or service) options on the website (which is in none of the academic measures but is in the practitioners' E-Ratings and Bizrate measures). Their list of component attributes for the *transaction* stage omits the start of the transaction. Important components omitted are ease of inputting personal details [15] and ease of payment, which can vary considerably even with a credit card as the usual mode of payment. Purchase attributes are essential components of the *transaction* stage, not the *website* stage even though they occur on the website, because consumers can interact with a website without purchasing.

5.3 Right First-Order Components (Items)?

As pointed out previously, Collier and Bienstock [5] failed to realize that the items they selected to represent the 11 "dimensions" themselves represent components: first-order components. According to C-OAR-SE, an item has an object part and an attribute part, and these first-order component attributes also have first-order component objects. For example, the items representing the attribute of (website) *design* (an attribute that is more correctly labeled as *design quality*; see the very good discussion of concrete cues versus perceptual attributes in [10]) have, as their respective first-order objects, *visual design* (items 10 and 11), *graphics* (item 12), *text* (item 13), and *page width* (item 14). The identification of first-order objects illuminates the mistake that Collier and Bienstock [5] made in conceptualizing each of the 11 "dimensions" as being reflective and then using factor analysis and stepwise deletion of items based on coefficient alpha. They dropped an unreported number of items (perhaps none?) from the five *website service quality* factors, three items (one each) from the three *transaction service quality* factors, and an astounding 31 items from the three factors representing *recovery service quality* (p. 268). Presumably, some or all of these deleted items were defining items, referring to first-order objects and attributes that make up the second-order attribute of *design quality*, and so forth, and are necessary for adequate content coverage of the construct. The defining items would also be necessary to form a valid diagnostic instrument for e-retailer practitioners.

5.4 Correct Procedure for Identifying Components and Generating Items

According to the C-OAR-SE procedure, for formed attributes, which the 11 "dimensions" should be, concrete (unambiguous) items should be generated from primary research consisting of qualitative interviews (semi-structured questions and open-ended answers) with a cross-section of e-retailers' prospective and actual customers. The items should then be categorized by two or three judges – experts in the domain of e-retailing – into first-order components, the main first-order components retained, and one content-valid item retained, or newly written if necessary, per first-order component (see [12], pp. 314-315). These items make up the scale. The next section suggests that there will be far fewer than the 54 items used by Collier and Bienstock [5].

6 Content of the Question Part and the Answer Part of the Items

6.1 Item Wording

Item wording (the question part of the item) is a fundamental yet much neglected aspect of content validity (see [12], pp. 320-324). Reviewers of new scales typically focus on scale statistics without bothering to examine the items on which the statistics are based. Researchers, in developing a new scale, typically seem to believe that a large number of often poorly worded items will somehow "cancel out" content errors, an approach adopted, apparently by Collier and Bienstock [5] in stating that "we have included numerous items for each [of the 11 'dimensions,' the second-order components] to ensure that content validity was achieved" (p. 269). The 54 questions in Collier and Bienstock's instrument, as well as the seven questions used to measure their two dependent variables, are reproduced in the present article as an Appendix (the questions are verbatim from the questionnaire in the Appendix to their article, pp. 272-273, but with new labels for the sets of questions). Inspection of the 54 items overall reveals that some of the questions present redundant content (e.g., items 6, 8 and 9; items 27 and 28; items 37, 39 and 43; and items 46 and 48). Redundancy of content has the effect of overweighting the first-order components that the items represent in the total score for the second-order component (the "dimension"). They also unnecessarily lengthen the instrument. Item 30 is logically incompatible with item 29 and should be dropped. The content suitability of the items for the separate service quality stages is assessed next, including these problematic items.

6.2 Website Service Quality

The content of the questions for the *website service quality* components (items 1 to 25) looks reasonably good, perhaps because most of the previous studies consulted by Collier and Bienstock [5] focused on website quality, and, most important, the specific (first-order) attributes in the questions do mostly relate to the formed (second-order)

attribute, *service quality*, such as *ease of navigation* (item 1), *security* (item 7), *readability* (item 13), *availability* (item 15), *objectivity* (item 20), and *site speed* (item 24).

6.3 Transaction Service Quality

At least four of the questions for *transaction service quality* (items 27, 28, 32 and 33) have inappropriate content. As pointed out earlier, these questions are about past transactions with the e-retailer instead of being about the most recent transaction, to which the other questions in the instrument refer (and which was the basis for the sampling of events, or transaction objects, in Collier and Bienstock's test of the instrument [5]). These questions assume that there have been past transactions and that the respondent is a repeat customer (the questions could not be meaningfully answered by first-time, and thus single transaction, customers). Use of the wrong attitude object, past transactions, has the likely result of nonsensical middle-of-the-scale "neither" answers for these attributes. Moreover, all of the items refer to products bought online and do not apply to online services, especially services purchased where there is no physical product other than the service receipt, such as e-tickets for air travel.

6.4 Recovery Service Quality

The 20 questions for *recovery service quality* (items 35 to 54) are almost surely too many, and this is after deleting 31 items due to inappropriate use of the "reflective" attribute operationalization. Ten items alone were used to measure the *recovery personnel competence* (Collier and Bienstock's [5] *interactive fairness*) component and it is difficult to accept that each is an important defining item. Qualitative interviews and expert judgment could have been used to identify the really highly important first-order components of the *recovery personnel competence* component and it is likely there would be no more than five or six. On the plus side, the items seem to relate to the overall quality attribute but to label this attribute as fairness, as Collier and Bienstock did, is dubious. On the other hand, the underlying attribute for *recovery outcome fairness* (Collier and Bienstock's *outcome fairness*) component is arguably fairness, not quality, but fairness must be convincingly argued to be a sub-attribute of service quality, otherwise the scores cannot meaningfully be summed to form the total *recovery quality* score (see [9] for the theoretical argument that not only outcome fairness but also customer-favored equity of the outcome are necessary components of service quality). However, there is still a content problem with the four questions (items 45 to 48) for the *recovery outcome fairness* component because one of them is a summary evaluation item rather than a concrete first-order component item (item 46: "The outcome I received was fair"). Factor-analysts have long warned about the inclusion of summary items among specific items because the summary item will almost always "attract" specific items to form a factor (e.g., [7]). The same problem occurs among the six items measuring *recovery procedure quality* because item 54 asks for a summary evaluation: "Overall, the e-retailer had a good procedure for dealing with complaints." Questions 46 and 54 should be dropped, further shortening the instrument.

6.5 Likert Answer Scales

Moreover, Collier and Bienstock's [5] use of Likert answer scales, which are advised against in C-OAR-SE, means that the attribute part of the item is built into the question part instead of the answer part (see [12], pp. 322-324). For example, consider item 3: "This e-retailer contains a site map with links to everything on the site." The first-order object here is *a site map with links to everything on the site* and the first-order attribute is *contains*. This question calls for a "yes" or "no" answer; it cannot be properly answered by indicating Likert-type "degrees of disagreement or disagreement." An example of a needlessly complicated and ambiguous item is item 6: "I trust the Website administrators will not misuse my personal information." The object in this item is the *administrators of the website* and the attribute could be read as double-barreled, involving whether the rater *trusts* the administrator and whether the administrator will *misuse* personal information. A conceptually simpler version, though still in the unsuitable Likert question-and-answer format, would be "The Website administrators are likely to misuse my personal information," "strongly disagree… strongly agree."

6.6 Correct Item Format

According to the C-OAR-SE procedure ([12], p. 323), the items should have a question part that identifies only the object, and then an answer part whose answer categories represent degrees of the attribute. In the case of *e-retailing service quality* items, the degrees of the attribute should represent a unipolar continuum. Qualitative research with consumers is necessary to specify the answer categories correctly in terms of how many and what categories of the attribute the raters can meaningfully distinguish. An example of the correct question-and-answer format for item 11 would be:

This e-retailer's Website design is:

□ not at all innovative □ moderately innovative □ very innovative

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 2 / ISSUE 3 / DECEMBER 2007 / 36 - 48
© 2007 Universidad de Talca - Chile

This paper is Available online at
www.jtaer.com

7 Enumeration (Scoring) to Improve the Predictive Validity and Managerial Usefulness of the Measure

7.1 Ambiguity of Likert Ratings

As well as being theoretically ambiguous, the Likert answer format is practically useless for managers. Firstly, to use item 11 again as an example, "This e-retailer's website design is innovative," how is the manager supposed to interpret, for instance, a mean rating of 3.5 on the 1-to-5 Likert answer scale that Collier and Bienstock [5] employed for innovativeness of design when the answer scale asks for agreement-disagreement and is bipolar? (The 3.5 here means .5 on the positive side of the 3.0 midpoint, whereas Collier and Bienstock used a reversed answer scale and would report this rating as 2.5, a further point of confusion for managers.) Does it mean that a "slightly innovative" design should be sought, in keeping with a bipolar interpretation of the answer scale, that is, only a slight departure from the conventional, expected design? Or do raters interpret the midpoint of the answer scale in a unipolar manner, so that 3.5 means "moderately innovative"? If the latter, the manager would want a design that is considerably more than "slightly innovative," which was the bipolar interpretation. Likert answer scales are ambiguous and cannot provide proper diagnostic information.

The problem with the Likert answer format can be solved, as recommended in the C-OAR-SE procedure, by abandoning the Likert question-and-answer format and instead using simple question wording referring to the object, with the answer wording referring to the attribute, and using meaningful, labeled answer categories. The correct format was illustrated earlier for item 11. The other items should be formatted the same way.

7.2 Enumeration (Scoring)

Enumeration, which is the final step in the C-OAR-SE procedure, is a conceptual step, not a mechanical, numerical formality in the application of the scale (see [12], pp. 324-326). The three components of *e-retailing service quality*, which are *website service quality*, *transaction service quality*, and *recovery service quality*, should be scored, and used, only as three separate constructs. The three stage quality scores should be used as sequential predictors: the *website service quality* score should be used to predict whether or not consumers proceed to a transaction (see [15]), and the *transaction service quality* score should be used to predict whether or not consumers will experience a problem that will engage the *recovery* stage. For those consumers who proceeded to a transaction, *website service quality* and *transaction service quality* can be used as multiple predictors, and *recovery service quality* used as an additional predictor for those who complained, to predict *future transaction intention*. Discriminant analysis, binary probit regression or, probably the best technique, the "equity estimator" [11], would be appropriate for testing these predictions.

7.3 Problems with the Other Constructs

In this connection, it should be pointed out that the *satisfaction* construct (just to label the attribute rather than the total construct) that Collier and Bienstock [5] included as a dependent variable was over-measured, with four redundant items (items 55 to 58 in the present Appendix table) used to produce a high and unnecessary coefficient alpha. Why the researchers would want to measure *satisfaction* in the first place is questionable given that this variable is, in effect, just a proxy for *overall service quality* ([14], p. 63). This is evidenced by the single summary item, item 56, "In general, I was pleased with the quality of service this e-retailer provided." Not only was there no need for an *overall service quality* variable but its inclusion in the regressions could have contributed to the lack of prediction exhibited by *transaction service quality* and *recovery service quality* by "suppressing" these two variables in predicting *behavioral intentions*.

Considering finally Collier and Bienstock's ultimate dependent variable, *behavioral intentions*, this construct should be reconceptualized as the consumer's *future transaction intention*, singular. The only relevant item of the three that Collier and Bienstock used is item 61: "I intend to purchase from this e-retailer in the future." The item before it, item 60, "I intend to continue to visit this e-retailer's site in the future," taps behavioral intention for only the first stage of e-retailing, which is a visit to the website. The item before that, item 59, "I will recommend this e-retailer to my friends," represents a different behavior altogether and should be analyzed separately (incidentally, three of the five items in Parasuraman, Zeithaml, and Malhotra's *loyalty intentions* measure [10], p. 231, refer to recommendations and thus seriously overweight this different behavior). That WOM recommendations are a different behavior is shown by the scoring of word-of-mouth intention across individuals as net (positive WOM minus negative WOM) which appears to provide a unique predictor, unlike other intended behaviors, of future sales growth [8]. Conceptually, the three items should not have been summed into a single score and, practically, the composite score is meaningless. The answer scale for *future transaction intention* should be one that can be enumerated into probabilities, such as the well-known Juster scale or Wallsten, Budescu, and Zwick's [17] scale as recommended in C-OAR-SE ([12], pp. 323-324), which is "impossible" (0), "unlikely" (.15), "slight chance" (.30), "toss up" (.50), "likely" or "good chance" (.70), "pretty sure" (.80), or "certain" (1.00). The insufficiently content-valid measure of intention may be another reason why Collier and Bienstock found the implausible results that *transaction service quality* failed to predict their dependent variable of

behavioral intentions and that *recovery service quality* also failed to predict *behavioral intentions* (see [5], Table 6, p. 270).

7.4 Several Domain Experts Should be Engaged to Develop Valid Measures

When implausible results like those found by Collier and Bienstock [5] emerge, service quality theorists should question the measures. In fact, the measures should be questioned rationally and a priori, before they are used, by examining them according to the guidelines in the C-OAR-SE procedure for scale development [12]. As explained in Rossiter ([12], see also [13]), designing or modifying measures so that they are judged by several experts in the domain of inquiry to have high content validity would prevent inappropriate tests of theories and hypothesized relationships and begin to counteract the many misleading empirical results that have been published in service quality studies and in management and marketing studies in general.

7.5 The Problem with New Measures Continues

After this critique was written, yet another new measure of e-retailing service quality appeared in the literature, called eTRANSQUAL [1]. The present author wishes to warn researchers that this other new scale exhibits all the problems that Collier and Bienstock's [5] does: although the items were commendably generated from original qualitative research, the items were wrongly assumed to represent eliciting attributes of service quality with the usual consequent deletion of defining items via the conventional procedure, and were wrongly measured with inappropriate Likert-type answer scales. Remarkably, although the researchers invested thought in identifying four plausible states of e-retailing (information search → agreement → fulfilment → after-sales service), they then ignored the stages in their measure and attempted, uselessly, to measure overall e-retailing service quality. Empirically, as did Collier and Bienstock, they wrongly sampled completed transactions. These researchers, too, could benefit from adopting the C-OAR-SE procedure to develop valid measures of the stages of e-retailing service quality for retailers engaged in electronic commerce.

References

[1] H.H. Bauer, T. Falk, and M. Hammerschmidt, eTransQual: A transaction process-based approach for computing service quality in online shopping, Journal of Business Research, vol. 59, no. 7, pp. 866-875, 2006.
[2] H.M. Blalock, H. M., Causal Inferences in Nonexperimental Research. Chapel Hill: University of North Carolina Press, 1964.
[3] K. Bollen and R. Lennox, Conventional wisdom on measurement: A structural equation perspective, Psychological Bulletin, vol. 110, no. pp. 2305-314, 1991.
[4] G.A. Churchill, Jr., A paradigm for developing better measures of marketing constructs, Journal of Marketing Research, vol. 16, no. 1, pp. 64-73, 1979.
[5] J.E. Collier and C.C. Bienstock: Measuring service quality in e-retailing, Journal of Service Research, vol. 8, no. 3, pp. 260-275, 2006.
[6] A. Diamantopoulos, The C-OAR-SE procedure for scale development in marketing: A comment, International Journal of Research in Marketing, vol. 22, no. 1, pp. 1-9, 2005.
[7] J.P. Guilford, J.P., Psychometric Methods. New York: McGraw-Hill, 1965.
[8] P. Kleinman, Philosopher's stone revealed?, Admap, May, p. 6, 2006.
[9] R.L. Oliver and J. E. Swan, Consumer perceptions of interpersonal equity and *satisfaction* in transactions: A field survey approach, Journal of Marketing, vol. 53, no. 2, pp. 21-35, 1989.
[10] A. Parasuraman, V.A. Zeithaml, and A. Malhotra, E-S-QUAL: A multiple-item scale for assessing electronic service quality, Journal of Service Research, vol. 7, no. 3, pp. 213-233, 2005.
[11] A. Rangaswamy and L. Krishnamurthi, Response function estimation using the equity estimator, Journal of Marketing Research, vol. 28, no. 1, pp. 72-83, 1991.
[12] J.R. Rossiter, The C-OAR-SE procedure for scale development in marketing, International Journal of Research in Marketing, vol. 19, no. 4, pp. 305-335, 2002.
[13] J.R. Rossiter, Reminder: A horse is a horse, International Journal of Research in Marketing, vol. 22, no. 1, pp. 23-25, 2005.
[14] R.T. Rust, A.J. Zahorik, and T.L.. Keiningham, Return on quality (ROQ): Making service quality financially accountable, Journal of Marketing, vol. 59, no. 2, pp. 58-70, 1995.
[15] C. Sismeiro and R.E. Bucklin, Modeling purchase behavior at an e-commerce website: A task-completion approach, Journal of Marketing Research, vol. 41, no. 3, pp. 306-323, 2004.
[16] J.H. Steckel, R.S. Winer, R.E. Bucklin, B.G.C. Dellaert, X. Dreze, G. Haubl, S.D. Jap, J.D.C. Little, T. Meyvis, A. L. Montgomery, and A. Rangaswamy, Choice in interactive environments, Marketing Letters, vol. 16, no. 3/4, pp. 309-320, 2005.
[17] T.S. Wallsten, D. V. Budescu, and R. Zwick, Comparing the calibration and coherence of numerical and verbal probability scales, Management Science, vol. 39, no. 2, pp. 176-190, 1993.
[18] V.A. Zeithaml, L. L. Berry, and A. Parasuraman, The behavioral consequences of service quality, Journal of Marketing, vol. 60, no. 2, pp. 31-46, 1996.

[19] Bryan Eisenberg, Jim Novo, John E. Shreeve (2006, May). The marketer's common sense guide to e-metrics [Online]. Available: www.futurenowinc.com/emetricsguide.pdf

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 2 / ISSUE 3 / DECEMBER 2007 / 36 - 48
© 2007 Universidad de Talca - Chile

This paper is Available online at
www.jtaer.com

Appendix

Collier and Bienstock's [5] Items in Their E-Service Quality Scale (Items 1-54) and Their Dependent Variables' Scales (Items 55-61). All the Questions are Answered on 5-Point Bipolar Likert Answer Scales (1 = Totally Agree, to 5 = Totally Disagree).

WEBSITE SERVICE QUALITY

Ease of Use

1. It is easy to get anywhere on this e-retailer's Website.
2. I don't get lost on this e-retailer's Website.
3. This e-retailer contains a site map with links to everything on the site.
4. This e-retailer's Website allows you to find a page previously viewed.
5. This e-retailer's Website allows you to go back when you make a mistake.

Privacy

6. I trust the Website administrators will not misuse my personal information.
7. Symbols and messages that signal the site is secure are present on this e-retailer's Website.
8. This e-retailer doesn't give other sites or companies access to my information.
9. This e-retailer doesn't give my information away to other companies.

Design Quality

10. This e-retailer's Website is visually pleasing.
11. This e-retailer's Website design is innovative.
12. I am able to see the graphics clearly on this e-retailer's Website.
13. This e-retailer's Website does not have fine print that is difficult to read.
14. I don't have to scroll from side to side to adequately see this e-retailer's Web page.

Information Trust

15. This e-retailer's advertised items are available in inventory.
16. This e-retailer provides information on how much an item costs with shipping costs included.
17. This e-retailer provides accurate information about when orders will be received.
18. This e-retailer's Website has a running total of purchases as the order progresses.
19. Prices are shown with the items on the screen.
20. This e-retailer's Website has information that is objective (i.e., product reviews are free from bias).

Functionality

21. When I use this e-retailer's Website there is very little waiting time between my actions and the Website's response.
22. This e-retailer's Website does not crash.
23. This e-retailer gives the customer numerous payment options.
24. This e-retailer's Website loads quickly.
25. The e-retailer provides a confirmation of items ordered.

TRANSACTION SERVICE QUALITY

Order Condition

26. This e-retailer's orders are protectively packaged when shipped.
27. All orders by this e-retailer are delivered undamaged.
28. Damage rarely occurs during transportation of my order from this e-retailer.

Timeliness

29. This e-retailer gives the customer multiple delivery time options (e.g., next day, 3- to 5-day delivery, or 5- to 7-day delivery).
30. The time between placing and receiving an order is short.
31. This e-retailer is able to respond to a rush order.

Order Accuracy

32. My orders from this e-retailer rarely contain the wrong items.
33. My orders from this e-retailer rarely contain incorrect quantities.
34. This e-retailer's billing is accurate.

Journal of Theoretical and Applied Electronic Commerce Research
ISSN 0718–1876 Electronic Version
VOL 2 / ISSUE 3 / DECEMBER 2007 / 36 - 48
© 2007 Universidad de Talca - Chile

This paper is Available online at
www.jtaer.com

Appendix (cont.)

RECOVERY SERVICE QUALITY

Interactive Fairness [Recovery Personnel Competence]

35. This e-retailer communicated honestly with me about my problem.
36. The e-retailer seemed very concerned about my problem.
37. I was given a reasonable explanation as to why the original problem occurred.
38. The e-retailer was courteous to me when trying to resolve my problem.
39. I believed what the e-retailer told me about how my problem occurred.
40. This e-retailer gives the customer the ability to talk to a “live” person using a telephone number.
41. The e-retailer was sympathetic and caring.
42. The e-retailer put a lot of positive energy into handling my problem.
43. The e-retailer told me why the service had failed in the first place.
44. The e-retailer was quite pleasant to deal with.

Outcome Fairness [Recovery Outcome Fairness]

45. Compensation was offered for problems the e-retailer created.
46. The outcome I received was fair.
47. In resolving my complaint the e-retailer gave me what I needed.
48. I got what I deserved.

Procedural Fairness [Recovery Procedure Quality]

49. The e-retailer responded quickly to my complaint.
50. The e-retailer adapted their complaint handling procedures to satisfy my needs.
51. I got a chance to tell the e-retailer the details of my problem.
52. The e-retailer showed flexibility in responding to my complaint.
53. The e-retailer made it easy for me to voice my complaint.
54. Overall, the e-retailer had a good procedure for dealing with complaints.

SATISFACTION

55. In general I (am/was) happy with the service experience.
56. In general, I was pleased with the quality of the service this e-retailer provided.
57. I was satisfied with the service this e-retailer provided.
58. I felt pretty negative about this e-retailer.

BEHAVIORAL INTENTIONS

59. I will recommend this e-retailer to my friends.
60. I intend to continue to visit this e-retailer’s site in the future.
61. I intend to purchase from this e-retailer in the future.

Copyright of Journal of Theoretical & Applied Electronic Commerce Research is the property of Universidad de Talca and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.