Available online at www.sciencedirect.com
ScienceDirect
Journal of Business Research 61 (2008) 903 – 911
JOURNAL OF BUSINESS RESEARCH

A comparison of importance weights and willingness-to-pay measures derived from choice-based conjoint, constant sum scales and best–worst scaling ☆

Jordan J. Louviere [a,1], Towhidul Islam [b,*]

[a] Centre for the Study of Choice (CenSoC), Faculty of Business, University of Technology, Sydney City Campus, Haymarket NSW 2090, Australia
[b] Department of Marketing and Consumer Studies, University of Guelph, Guelph, Canada ON, N1G 2W1

Received 4 March 2004; accepted 14 November 2006

Abstract

We review the measurement of product attribute importance, and find little consensus in definition or measurement methods. We compare four measurement methods: 1) two direct methods whereby respondents report the importance of attributes using best–worst scaling or constant sum scales, and 2) two indirect methods derived from discrete choice experiments. Our comparisons rely on previous findings that choice experiments are externally valid to use as the standard. We find high agreement within direct or indirect methods, but less agreement between direct and indirect methods. Our results also demonstrate that inferences derived from indirect measures appear to be susceptible to context effects related to the particular attributes a researcher chooses to investigate. We discuss implications for current and future research.

Keywords: Relative importance; Discrete choice experiments; Willingness-to-pay (WTP); Missing attribute information

1. Introduction

The concept of weight or importance in judgment and decision making has a long history, expressed in various ways not only in most social and business sciences but also engineering, physical sciences and medicine. Thus, while the notion is pervasive, there is little agreement within or across disciplines regarding the meaning of weight or importance, much less how to measure it. For example, is weight a measure of the attentional focus/concentration a person gives to a dimension, cue or piece of information in a decision making task or situation (e.g. Anderson, 1971)? Or, is weight a measure of the statistical impact that a dimension, cue or piece of information has in a particular task or situation (e.g. Green and Kriger, 1995)? Or, is weight non-existent, a quantity not mathematically separable from scale value/position of values/levels of dimensions/cues/information on an underlying scale, as implied in axiomatic utility theory and/or conjoint measurement (e.g. Keeney and Raiffa, 1976; Louviere, 1988)?

Over two decades ago, Shanteau (1980) tried to review the wide array of concepts that denoted or connoted "weight", including operational definitions and ways proposed or used to measure it. Shanteau concluded that there were so many definitions and concepts that there was little common ground. Not much has changed since then, except that many more academic and commercial applications of "weight" have appeared in many literatures.

Our aim is to make a small contribution to the measurement of weight or importance by comparing four ways to measure it. There are two general ways to measure weight or importance — directly or indirectly. Direct approaches typically try to measure the importance of a set of dimensions by asking individuals or groups to state the degree of importance or weight on some elicitation scale like a category rating scale or a constant sum scale. Indirect approaches vary

☆ The authors gratefully acknowledge funding provided by Social Sciences and Humanities Research Council (SSHRC) of Canada Grant # 046288 and University of Guelph Internal SSHRC Research Grant # 045415.
* Corresponding author. Tel.: +1 519 824 4120x53835; fax: +1 519 823 1964.
E-mail addresses: jordan.louviere@uts.edu.au (J.J. Louviere), islam@uoguelph.ca (T. Islam).
[1] Tel.: +61 2 9514 3993; fax: +61 2 9514 3535.

0148-2963/$ - see front matter
doi:10.1016/j.jbusres.2006.11.010

widely, but generally try to infer weight or importance by analyzing an outcome measure like choices. Typically, outcomes are expressed as some function of a set of dimensions, often as some form of generalized regression function. The function specified depends on the problem, which varies widely between disciplines.

We compare four approaches in this paper: 1) constant sum scaling of product attribute/feature importance (e.g. Churchill, 1986, p. 432; Lehman et al., 1998, pp. 247–248); 2) best–worst scaling of product attribute/feature importance (e.g. Finn and Louviere, 1993; Marley and Louviere, 2005; Flynn et al., 2007); 3) product attribute/feature importance inferred from statistical effects in discrete choice experiments (e.g. Louviere and Woodworth, 1983; Louviere et al., 2000); and 4) implied willingness-to-pay for differences in attribute levels that measures the compensating variation in one variable (i.e., price) required to offset a change in a second (e.g., a difference in attribute levels). To anticipate our results, we find positive associations among all four methods for delivered pizza products, but less agreement for packaged fruit juice products. Both categories exhibited high correlations between direct measures (best–worst, constant sum) and high correlations between indirect measures (relative statistical effect, willingness-to-pay), with lower correlations between direct and indirect measures.

2. Direct and indirect measures of weight/importance

Cohen and Neira (2003) and Cohen (2003) note that often there is little differentiation among attribute/feature importances measured on category rating scales. They show that responses to attribute/feature importance rating questions reflect little time on each rating, with most items rated as relatively important. Indeed, this motivated Finn and Louviere (1993) to develop the best–worst scaling approach (also called "maximum difference scaling") because it had a cognitive psychological basis, was easy to implement, was easy for respondents, and also encouraged respondents to trade off attributes/features. Finally, other authors like Srinivasan (1988) also find differences in indirect and direct measures, such as an indirect, conjoint-derived measure being superior to a direct measure. Surprisingly few other comparisons are available, with no comparisons of the four measures we study below.

Such cross-validations/comparisons matter because much consumer behavior research requires an ability to differentiate attribute/feature importance and/or differentiate attributes/features in other ways. For example, academic and applied researchers often use a priori qualitative research to identify attributes that consumers use to make decisions, then measuring attribute/feature importance to select a final set of attributes to study. If the methods that researchers use to do this give different conclusions about attribute/feature importance, there would be significant implications for academic and applied researchers. Hence, it is important to compare methods to determine if conclusions about attribute importance differ, which allows us to make a modest contribution to measuring attribute/feature weight or importance.

2.1. Attribute importance may be affected by context

Why do many disciplines think that humans can directly output measures of relative importance, or at least the order of importance of attributes/features? Indeed, this would seem to be a "big ask" of humans as the importance of an attribute like price should be context-dependent. That is, one might expect that such answers should depend on the ranges of price values one has previously experienced, expects to experience or are provided by researchers. If all subjects have the same frame of reference – say \$8 to \$12 for a small pizza – asking how important price is relative to other pizza attributes should be meaningful. However, if individuals have different reference frames whereby some believe the range of prices is \$12 to \$18, but others believe \$8 to \$12, it is unclear if one can compare importance measures from different individuals without knowing their reference frames.

Several researchers (e.g. Tversky and Simonson, 1993; Green and Kriger, 1995; Rohrbaugh and Shanteau, 1999) show that context affects derived attribute/feature importance. Our research manipulates presence/absence of attribute information in product descriptions; so we expect derived attribute/feature importance measures to be context-dependent because prior work shows tradeoffs and preferences are context-dependent (e.g. Meyer, 1981; Huber and McCann, 1982; Johnson, 1987; Levin et al., 1986; Lynch and Srull, 1982; Kivetz and Simonson, 2000; Simmons and Lynch, 1991; Johnson and Levin, 1985).

We vary presence/absence of attributes because academics and practitioners suggest limiting numbers of attributes in conjoint and choice experiments, implying a basis for excluding one or more attributes (e.g. Carson et al., 1994). To wit, various ways to limit numbers of attributes have been proposed like the partial profile approach of Chrzan and Elrod (1995) or hybrid conjoint proposed by Green (1984). Other researchers arbitrarily limit numbers of attributes by using ad hoc methods to identify "salient attributes", like qualitative approaches discussed by Louviere (1988) and Louviere et al. (2000).

2.2. Importance may be affected by ambiguity

Issues of context-dependent attribute/feature importance are similar to issues related to ambiguity effects associated with attributes/features in decision making tasks (see e.g., Louviere, 2001; Louviere et al., 2002). That is, if the color of a Kitchenaid mixer is a feature, but we do not know what colors individuals think about when asked how important color is, the distribution of importance is conditional on the distribution of ambiguity/uncertainty about the colors used as reference frames. Specifically, verbally describing mixer color levels as "black", "white" and "blue" may result in all individuals forming similar mental images of "black" and "white", but which shade(s) of "blue" individuals have in mind is unclear. So, if there is a distribution of imagined "blues" across individuals, preference estimates for mixer colors will be confounded with this distribution.

One needs to control for such confounds in studying decisions. If one uses direct measures of attribute importance, respondents