

Università degli Studi di Padova



Facoltà di Scienze Statistiche

Corso di Laurea Triennale in Statistica e Gestione delle Imprese

TESI DI LAUREA

**Le procedure di sviluppo di scale che misurano costrutti
di marketing: teorie ed alcuni esempi pratici**

**Procedures for scale development in marketing: theories and
some practical examples**

Relatore: Prof. Francesca Bassi

Laureando: Federico Faoro – 554051

ANNO ACCADEMICO 2008-09

INDICE

Introduzione	pag.4
Capitolo 1. La prima teoria strutturata sullo sviluppo di scale di misura	pag.6
1.1	Introduzione
1.2	La misura della soddisfazione dei venditori industriali
1.2.1	La realizzazione della scala
1.3	La teoria di Churchill
1.3.1	Affidabilità e validità
1.3.2	La procedura di sviluppo della scala
1.4	Lo studio di Peter sull'affidabilità
1.4.1	Approcci tradizionali all'affidabilità
1.4.2	Creazione di una teoria generalizzata
1.4.3	Affidabilità in precedenti studi di marketing
Capitolo 2. Evoluzione e sviluppo delle teorie	pag.19
2.1	Introduzione
2.2	Lo studio di variabili unidimensionali di Anderson
2.2.1	Gli elementi distintivi della teoria: unidimensionalità e analisi fattoriale confermativa
2.2.2	Confronto tra analisi fattoriale confermativa e altri metodi
2.3	Un'alternativa alle scale di misura: gli indicatori formativi
2.3.1	Un approccio teorico
2.3.2	La costruzione dell'indice
2.4	La procedura C-OAR-SE
2.4.1	L'impianto teorico della procedura C-OAR-SE
2.4.2	Accertamento delle proprietà di affidabilità e validità
2.5	La teoria di generalizzabilità multivariata
2.5.1	Rassegna storica delle procedure di sviluppo di scale di misura
2.5.2	L'approccio della teoria di generalizzabilità multivariata
2.5.3	Il modello matematico alla base della teoria di generalizzabilità multivariata
2.5.4	Un esempio pratico

Capitolo 3. Studio di alcune variabili di marketing

pag.36

- 3.1 Introduzione
- 3.2 Il coinvolgimento del consumatore
 - 3.2.1 Il concetto di coinvolgimento
 - 3.2.2 Il processo di sviluppo della misura
- 3.3 I taboo nella pubblicità
 - 3.3.1 Il concetto di taboo
 - 3.3.2 I taboo e la pubblicità
 - 3.3.3 Lo sviluppo della scala
 - 3.3.4 Effetti dell'uso di taboo nella pubblicità
- 3.4 L'orientamento al mercato di un'impresa
 - 3.4.1 La realizzazione della scala

Bibliografia

pag.50

Introduzione

Le scale di misura degli atteggiamenti costituiscono uno strumento indispensabile per gli operatori nell'ambito del marketing al fine di conseguire informazioni che altrimenti non sarebbero ottenibili. Le variabili di marketing che sono di cruciale interesse per le aziende infatti non sono direttamente osservabili in quanto esse rappresentano degli atteggiamenti o delle attitudini di un determinato soggetto: si pensi ad esempio alla soddisfazione di un cliente nei confronti di un prodotto, oppure al coinvolgimento di una certa pubblicità o ancora alla fedeltà ad una marca.

Da un punto di vista storico il problema dello sviluppo di scale di misura ha visto fino alla fine degli anni settanta la creazione di strumenti utili solamente a determinate situazioni che, da un lato, non consentono ovviamente la possibilità di generalizzare ad altri ambiti e, dall'altro, non possiedono quelle proprietà fondamentali di bontà che una scala deve possedere.

Nel 1979 si assiste tuttavia ad una svolta anche sull'onda dell'importanza sempre maggiore che il marketing stava ottenendo all'interno dell'attività di un'impresa. Sul *Journal of Marketing Research* appaiono infatti due articoli fondamentali che forniscono un'ampia e solida teoria strutturata sullo sviluppo di scale di misura: "A Paradigm for Developing Better Measures of Marketing Constructs" di Gilbert A. Churchill Jr e "Reliability: A Review of Psychometric Basics and Recent Marketing Practices" di J. P. Peter. L'importanza di questi due lavori sta nel fatto che essi rappresentano e forniscono un vero e proprio modello a cui affidarsi se ci si trova davanti al problema della costruzione di una scala di misura. In particolare, il primo descrive un percorso operativo da seguire e il secondo si focalizza sulle proprietà che un buono strumento deve possedere.

Da quel momento in poi, le idee e i risultati di Peter e Churchill vengono ampiamente sfruttati a tal punto che anche a livello teorico essi non subiscono sostanziali modifiche se non alcuni sforzi di innovazione quale quello effettuato nel 1988 da Anderson con l'articolo "An Update Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment", dove il carattere nuovo è rappresentato dall'introduzione di uno strumento (l'analisi fattoriale confermativa) utile in alcuni contesti di ricerca.

Di rilievo è invece uno studio che compare nel 2002 firmato da Rossiter dal titolo "The C-OAR-SE procedure for scale development in marketing": sulle orme di Churchill, ma con un approccio diverso, viene esposta una nuova teoria alternativa di sviluppo di scale di misura articolata anch'essa su una serie di stadi da seguire, caratterizzata da un'attenta e precisa definizione del costrutto di marketing in analisi.

Infine in tempi ancora più recenti compare “*How fine is C-OAR-SE? A generalizability theory perspective on Rossiter’s procedure*”(2005) scritto da A. Finn U. Kayand e pubblicato sull’*International Journal of Marketing*, un articolo chiave che cerca di unire gli aspetti positivi delle tesi di Churchill e Rossiter superando i rispettivi limiti con l’ottica di conseguire una migliore procedura di sviluppo di scale di misura.

Questo lavoro ha dunque lo scopo di presentare le varie acquisizioni a cui si è arrivati nel corso del tempo in materia di sviluppo di misure, non solo descrivendo i vari articoli sopra citati di natura fondamentalmente teorica, ma anche mostrando applicazioni pratiche a costrutti di marketing quali il coinvolgimento del consumatore, l’orientamento al mercato di un’azienda e il taboo nella pubblicità.

Capitolo 1: La prima teoria strutturata sullo sviluppo di scale di misura.

1.1 Introduzione

“*A Paradigm for Developing Better Measures of Marketing Constructs*” di Gilbert A. Churchill Jr e “*Reliability: A Review of Psychometric Basics and Recent Marketing Practices*” di J. P. Peter, entrambi pubblicati sullo stesso numero del *Journal of Marketing Research* nel 1979, rappresentano senza dubbio due pietre miliari nell’ambito degli studi sullo sviluppo di scale di misura.

Degno di rilevanza è però anche il precedente “*Measuring the Job Satisfaction of Industrial Salesmen*” (1974) di Churchill, nel quale sono ben evidenti i semi della teoria churchilliana che verrà appunto alla luce cinque anni più tardi.

1.2 La misura della soddisfazione dei venditori industriali

L’articolo “*Measuring the Job Satisfaction of Industrial Salesmen*” (1974) comparso sul *Journal of Marketing Research* offre utili ed interessanti spunti di riflessione e rappresenta un buon punto di partenza nella narrazione delle varie risposte che vengono date al problema dello sviluppo di scale per la misura degli atteggiamenti.

Esso tratta della misura di un costrutto di marketing, ovvero il livello di soddisfazione dei venditori del settore dell’industria; è scritto da Churchill assieme a Ford e Walker e, soprattutto, è antecedente al fondamentale studio sulla procedura di sviluppo di scale di misura che verrà condotto dallo stesso Churchill, essendo stato pubblicato cinque anni prima.

Risulta quindi molto interessante studiare l’approccio riservato nella trattazione della costruzione di uno strumento di misura del costrutto per evidenziare somiglianze e discordanze con quanto verrà teorizzato pochi anni dopo dallo stesso autore.

La focalizzazione su questa variabile è giustificata dal fatto che il ruolo del venditore è estremamente cruciale e delicato. Esso rappresenta infatti il soggetto attraverso il quale un’azienda si relaziona con il mercato e allo stesso tempo costituisce un fattore decisivo nel successo di un’impresa.

Ecco quindi che si rivela necessario misurare quanto un venditore sia soddisfatto del proprio lavoro e capire quali siano i fattori alla base di questo fenomeno. In generale si vuole, inoltre, creare uno strumento adeguato ed appropriato per descrivere la variabile, visto che la letteratura fino a quel momento aveva prodotto scale inefficaci non in grado di cogliere tutte le sfumature e le componenti.

1.2.1 La realizzazione della scala

Il primo passo da compiere è quello di stabilire il dominio di ricerca identificando le componenti che costituiscono il costrutto preso in esame.

Oltre a fare riferimento a precedenti studi, è stato necessario condurre delle interviste a domande aperte coinvolgendo una serie di venditori nel campo dell'industria e di psicologi che hanno avuto tra i loro pazienti proprio dei venditori. Alla fine, sono state individuate otto componenti che sono alla base del livello di soddisfazione: il lavoro stesso, il rapporto con i colleghi, la supervisione, la formazione alle vendite e il supporto a casa, l'azienda e i suoi dirigenti, la paga e gli incentivi, la carriera e i clienti.

Sono stati poi selezionati gli item per un totale di 185, facendo riferimento ancora una volta a quanto presente in letteratura e alle interviste condotte. È stata inoltre utilizzata la scala di Likert e, per evitare distorsioni nei risultati, gli item sono stati messi in ordine casuale e metà di essi sono stati formulati negativamente.

Una prima somministrazione ad un campione di venditori di un settore manifatturiero è stata condotta per purificare la misura ed escludere alcuni item attraverso il calcolo della correlazione tra il punteggio dei singoli item con il punteggio totale e del coefficiente alfa di Cronbach; alla fine il numero di item è sceso a 117. In questa fase si è ricorso al *domain sampling model*, un modello che prevede la selezione casuale e il conseguente utilizzo di una parte degli item concepiti.

Vista la modifica della scala e visto il campione precedente preso da una singola azienda è stata eseguita una nuova somministrazione. Gli item sono stati preliminarmente divisi in due gruppi a seconda della loro correlazione con il punteggio totale: item con simile correlazione vengono assegnati casualmente a uno dei gruppi. Il campione, costituito da 479 venditori, è stato scelto tra sette diverse aziende ed è stato diviso in due gruppi ai quali sono stati somministrati a distanza di circa due settimane ad ordine inverso le due parti del questionario. La prova dell'affidabilità attraverso questa divisione in due blocchi è stata ottenuta grazie alla significativa correlazione tra i due punteggi che è stata riscontrata.

È infine stata ricercata un'eventuale ambiguità nella formulazione degli item inviando ai rispondenti un ulteriore questionario.

La realizzazione dello strumento non è tuttavia ancora finita in quanto è necessaria un'analisi fattoriale per capire se l'individuazione delle otto componenti iniziali è corretta.

Oltre ad eliminare 22 item che sono risultati poco incidenti sulle rispettive componenti, attraverso questa analisi, è stata confermata la presenza di 5 componenti (il lavoro, i colleghi, la carriera, i clienti) e dalle tre rimanenti sono stati ricavati "paga" e "supporto e politica dell'impresa" per un totale finale di sette componenti.

Effettuata questa ulteriore selezione degli item bisogna ricorrere al calcolo del coefficiente alfa di Cronbach sia per ciascuna componente che per l'intero questionario per accertarsi della proprietà di affidabilità. La scala è stata anche divisa in due parti somministrate allo stesso campione in due tempi diversi per prevenire errori che non dipendono dallo strumento ma appunto dalle situazioni e condizioni di somministrazione calcolando in aggiunta la correlazione tra stesse componenti delle due metà.

I risultati di questo procedimento hanno prodotto indici di correlazione e coefficienti alfa sufficientemente alti per stabilire l'affidabilità dello strumento.

L'altra proprietà indispensabile che una scala di misura deve avere è quella della validità ed in particolare in questa situazione la validità del costrutto.

Questa si ottiene innanzitutto eseguendo una corretta specificazione del dominio senza la quale si rischia di compromettere già in partenza la qualità della misura. Anche la coerenza interna in questa fase è molto informativa in quanto ci dice se tutti gli item che utilizziamo colgono lo stesso costrutto. Ma la validità del costrutto ha a che fare anche e soprattutto con il saper cogliere attraverso la scala ciò che realmente ci si è prefissati in partenza e i due aspetti appena citati non sono sufficienti. Per questo motivo Churchill suggerisce di studiare la relazione di questo costrutto con altri quali, ad esempio, la performance, tutto a patto che la teoria fornisca in partenza prove certe del legame: una significativa correlazione deporrebbe in favore della validità cercata.

L'ultimo passo da compiere dopo aver raggiunto la versione definitiva della scala a 95 item è quello di stabilire delle norme per l'interpretazione dei punteggi che allo stato grezzo risulterebbero poco informativi.

Ciò che si effettua qui è una standardizzazione e una normalizzazione attraverso la media e la deviazione standard dei punteggi ottenuti; in questo modo i dati sono interpretabili e utili per trarre opportune conclusioni.

In questo articolo, Churchill spiega attraverso un percorso ben delineato e chiaro come ottenere una scala che va a misurare un costrutto molto importante che fino ad allora non aveva visto la realizzazione di scale altrettanto solide dal punto di vista della costruzione stessa e della verifica delle proprietà di affidabilità e validità.

Aspetto altrettanto importante è il fatto che Churchill non fa altro che anticipare quanto pubblicherà qualche anno più tardi: il processo qui descritto per realizzare la scala coincide sostanzialmente con quello che egli avrà modo di teorizzare più avanti. La differenza principale che si può riscontrare tra i due lavori di Churchill, che sarà più chiara con la descrizione del secondo nel prossimo paragrafo, resta semplicemente il loro carattere evidentemente dissonante: il primo infatti è più che altro applicativo senza pretese di venire generalizzato in altri ambiti, il secondo invece, come si vedrà, è ideato proprio per fornire degli strumenti da utilizzare in molteplici situazioni.

1.3 La teoria di Churchill

Il primo studio sistematico di procedure per lo sviluppo di scale di misurazione di variabili di marketing avviene col fondamentale articolo di Gilbert A. Churchill Jr, “*A Paradigm for Developing Better Measures of Marketing Constructs*”, pubblicato nel febbraio del 1979 nel *Journal of Marketing Research*.

Prima di allora infatti nessuno studioso del settore si era occupato in maniera precisa della materia e per questo motivo non esisteva una teoria di riferimento per chi si avvicinava ai problemi di misura. Agli operatori di marketing quindi non restava che affidarsi ciecamente a quella scarsa e poco organizzata letteratura presente sull’argomento senza tuttavia adottare uno spirito critico. Il bisogno di migliori misure c’era ma ciò che non c’era erano gli strumenti per ottenerle.

Ecco quindi che la missione di Churchill risulta essere quella di fornire una teoria strutturata sullo sviluppo di scale di misura, puntando in particolare l’attenzione sulle proprietà di affidabilità e validità che una scala deve possedere.

1.3.1 Affidabilità e validità

Prima di affrontare la vera e propria descrizione del processo attraverso il quale si sviluppa una scala di misura è indispensabile chiarire alcuni concetti fondamentali.

Nel misurare il livello di un costrutto di marketing quale ad esempio può essere la customer satisfaction gli studiosi si attendono un certo livello X_t , a fronte di un livello X_o effettivamente ottenuto tramite la misurazione.

I due livelli teoricamente dovrebbero coincidere, tuttavia possono sorgere delle differenze dovute sostanzialmente ad errori di due tipi, ovvero errori sistematici ed errori casuali. I primi si riferiscono a vere differenze riscontrabili nella natura di ciò che si cerca di misurare, i secondi invece sono legati ai soggetti valutatori (ovvero le persone intervistate, la cui predisposizione a rispondere o il cui stato d'animo possono avere delle influenze), alle condizioni in cui vengono somministrate le scale (a casa, in un centro commerciale, ...), al campionamento inadeguato o all'ambiguità degli item, a semplici errori di segnatura di risposta o ad errori di codifica.

Tutti questi errori hanno dunque un certo impatto sul livello finale misurato, un impatto che comunque può essere previsto tramite la seguente relazione funzionale:

$$X_o = X_t + X_s + X_r,$$

dove X_s rappresenta le fonti sistematiche di errore dovute alla natura dell'oggetto che si misura e le fonti casuali di errore.

A fronte di questa relazione si perviene a due concetti fondamentali associati ad una scala di misura, che sostanzialmente ne implicano e certificano la qualità: il concetto di validità e il concetto di affidabilità.

Una misura è valida quando $X_o=X_t$, ovvero quando le uniche differenze riscontrabili tra i livelli (atteso e misurato) sono dovute a reali differenze nelle caratteristiche di ciò che si sta misurando.

Una misura è invece affidabile se produce risultati concordanti con un'altra misura indipendente dalla prima ma comparabile dello stesso oggetto che si vuole misurare, ovvero se gli errori casuali vengono annullati ($X_r=0$).

Tutto ciò implica, come si può osservare dalla relazione funzionale, che l'affidabilità è una condizione necessaria ma non sufficiente per la validità, cioè che se una misura è valida è anche affidabile, ma non viceversa.

Risulta evidente come il principale obiettivo sia quello di ottenere un valore X_o il più possibile vicino ad X_t . Contemporaneamente però ci si scontra con un altro problema: non si può mai sapere esattamente il valore di X_t , e quindi occorre fare inferenza per ottenere delle stime servendosi di indici di affidabilità e di validità (che verranno descritti in seguito).

Uno dei vincoli principali del processo di sviluppo di scale di misura proposto da Gilbert e Churchill è la necessità di utilizzare più di un item per descrivere un costrutto o ciascuna delle diverse dimensioni che esso stesso possiede.

I motivi di questa scelta sono sostanzialmente tre e sono dovuti al fatto che,utilizzando singoli item si otterrebbero: bassa correlazione con l'attributo oggetto di studio, data la specificità risultante da una singola domanda e quindi la sua inadeguatezza a cogliere le differenze nell'oggetto di studio; insufficiente "segmentazione" dei soggetti rispondenti,ovvero scarsa categorizzazione delle persone; errori di misura elevati che rendono la misura poco affidabile.

In sostanza, un singolo item viene considerato ben lontano dal descrivere in maniera precisa il costrutto oggetto di studio e la scelta di utilizzare più item viene ritenuta in grado di garantire affidabilità e validità.

1.3.2 La procedura di sviluppo della scala

Veniamo ora alla descrizione del processo di sviluppo di scale di misura. Churchill e Gilbert propongono nel loro articolo una sequenza di passaggi, otto in totale, ognuno dei quali supportato da alcune operazioni da svolgere,e consistono in: specificazione del dominio del costrutto, generazione di un campione di item, raccolta dati, purificazione della misura, raccolta di nuovi dati, accertamento dell'affidabilità con i nuovi dati, accertamento della validità del costrutto, sviluppo di norme.

- *Specificazione del dominio del costrutto*

Il primo passo da compiere è quello di delineare correttamente il dominio della variabile di marketing oggetto di studio, ovvero quello di darle una precisa definizione cercando di catturare tutte le sue componenti o dimensioni.

È importante delineare i confini entro i quali si svolge lo studio del ricercatore,capire cosa incorporare e cosa lasciare fuori, facendo magari riferimento, tra le altre cose, alla letteratura presente sull'argomento oggetto di ricerca.

- *Generazione di un campione di item*

Una volta specificato il dominio entro il quale si opera, la procedura prevede la scelta di un campione di item che riesca a cogliere a pieno il dominio appena delineato.

Anche in questo caso si può far riferimento alla letteratura preesistente sulla variabile di interesse, ma si può anche ricorrere a sondaggi tra soggetti esperti o in qualche modo coinvolti (come i consumatori) che possano fornire idee e spunti tramite ad esempio dei focus group.

Nella realizzazione di ogni singolo item bisogna poi cercare di usare la stesura più adeguata per esprimere in maniera precisa il concetto che si vuole esporre, evitando ad esempio ambiguità nella formulazione. Una volta scelti gli opportuni item si passa alla fase di raccolta dei dati.

- *Purificazione della misura*

Una volta acquisiti i dati si procede coll'effettuare una serie di calcoli per purificare la misura, ovvero per verificare se esiste quella che viene chiamata coerenza interna: si ricorre ad indici di correlazione, al coefficiente alfa ed anche all'analisi fattoriale.

In questa fase si fa riferimento al *domain sampling model*, un modello che consiste nel somministrare non tutti gli item ma una parte di essi che siano comunque rappresentativi al fine di ottenere un punteggio X_o il più possibile prossimo al vero livello X_t .

Gli indici a cui si è interessati sono la correlazione tra gli item, la correlazione media che prova quanta somiglianza abbiano tra loro tutti gli item e un indice di dispersione sulla correlazione media che mostra fino a che punto gli item variano. È ovvio che se si suppone che gli item appartengano allo stesso dominio ci si deve aspettare una correlazione relativamente alta; in caso contrario è tuttavia necessario rivalutare la definizione del dominio.

Un altro indice molto utile per attestare la bontà dello strumento che si calcola in questa fase è il coefficiente alfa di Cronbach, la cui radice quadrata corrisponde alla correlazione stimata di k-item con il vero punteggio (X_t). Un alfa basso indica infatti lo scarso legame dei k-item con il dominio della variabile oggetto di studio, e per questo motivo si tende ad eliminare tutti quei gruppi di item ai quali è associato un coefficiente relativamente basso prossimo allo zero.

Un ulteriore studio dei dati viene fatto attraverso l'analisi fattoriale. Con questo strumento si cerca di ottenere delle prove empiriche che confermino le diverse dimensioni della variabile di interesse concettualizzate nelle fasi precedenti.

Una volta condotte queste analisi si procede col valutare la qualità dei risultati: se si sono ottenuti esiti positivi (alfa alto, dimensioni del costrutto ben individuate) si passa alla raccolta di nuovi dati; in caso contrario bisogna invece ripercorrere le fasi precedenti cercando di individuare ed evitare gli errori che hanno portato a determinare una scala inappropriata.

- *Accertamento dell'affidabilità con nuovi dati*

Il *domain sampling model* descritto precedentemente può rappresentare uno strumento buono per ridurre gli errori casuali, responsabili di influenzare la qualità di una scala.

Alcuni errori di questa natura infatti (come ad esempio la presenza di item vaghi od ambigui) tendono ad abbassare la correlazione tra item come pure il coefficiente alfa: in presenza di questi valori bassi non si fa quindi che eliminare gli item in questione, mantenendo in tal modo l'affidabilità della misura.

Tuttavia il coefficiente alfa non è utile nel verificare l'affidabilità nei casi in cui le fonti di errore non dipendono dallo strumento ma da fattori ambientali o personali legati all'intervistato. In questo caso è allora necessario raccogliere nuovi dati, e solo dopo fare le dovute considerazioni sull'affidabilità dello strumento.

- *Accertamento della validità del costrutto*

Passo dopo passo si è arrivati a costruire una misura valida nei contenuti ed affidabile, i cui item si dice possiedano quella che viene chiamata coerenza interna.

Tuttavia non si può dire di avere ottenuto a questo punto un'altra proprietà importante, ovvero la proprietà di validità del costrutto; questo concetto è direttamente legato alla natura di ciò che si sta misurando, e la coerenza interna conseguita è necessaria ma non sufficiente a determinare la validità del costrutto.

Per verificare la validità si deve allora studiare la correlazione tra due diverse misure che cercano di studiare la stessa variabile e si deve anche capire se la misura stessa si comporta come ci si attendeva si comportasse.

Quando si va a misurare un fenomeno è sempre preferibile avere almeno due strumenti di misurazione: i vantaggi sono ovvi, in quanto un singolo strumento non può garantire risultati attendibili nella definizione della natura dell'oggetto che si vuole studiare.

Dal confronto delle misurazioni di due scale diverse si possono trovare elementi in comune ed elementi distintivi: si parla in questi casi rispettivamente di validità convergente e validità discriminante.

La validità convergente si ha quando vi è una forte correlazione tra le due diverse misure: in questo caso si ha la prova che si sta misurando la stessa variabile e che questa variabile non è semplicemente un artefatto frutto della ricerca di uno studioso bensì un elemento reale.

La validità discriminante è invece auspicabile quando si sta cercando di creare una nuova scala migliore delle precedenti: una correlazione non troppo forte tra le misure testimonia infatti una

presenza di qualche nuovo elemento che è stato introdotto, ovvero la nuova scala non è mera copia di quella precedente.

Per studiare questi due tipi di validità si può ricorrere alla cosiddetta “multitrait - multimethod matrix” (utilizzabile anche nell’ambito di analisi fattoriali) i cui elementi rappresentano la correlazione tra dimensioni di una variabile misurati da differenti metodi.

Per accertare la validità si pone l’attenzione tra le altre cose sui valori della correlazione tra i punteggi che si ottengono somministrando due diverse scale per misura la stessa componente di un costrutto (elementi sulla diagonale): valori significativamente diversi da zero sono una buona prova della validità che si va cercando.

Il tutto tenendo ovviamente in considerazione il fatto che sia le componenti della variabile oggetto di studio sia i due metodi di misura devono essere il più indipendenti possibile; in caso contrario ci si scontrerebbe con una correlazione troppo alta che allontanerebbe il ricercatore dall’obiettivo di ottenere da una parte la validità sperata e dall’altra una misura più efficace.

A questo punto del processo di sviluppo di una scala di misura si propone un ulteriore interrogativo. La scala fin qui elaborata gode delle suddette proprietà, tuttavia bisogna ancora verificare se ciò che essa misura è effettivamente ciò che il ricercatore si è proposto di studiare, ovvero si cerca una verifica risolutiva della validità del costrutto.

Per risolvere questo problema si introduce allora una seconda variabile che si assume abbia dei legami accertati con quella di partenza e una sua misura valida ed affidabile: se si ottengono due scale positivamente correlate si ha allora ottenuto la definitiva prova della validità del costrutto.

- *Sviluppo di norme*

I punteggi di una scala di misura, presi singolarmente, sono poco informativi. Ad esempio un singolo livello di una misura della soddisfazione nei confronti di un prodotto non può dare una chiara risposta sul livello di soddisfazione del soggetto. Per questo motivo è necessario sviluppare delle norme, ovvero bisogna ad esempio porre a confronto i diversi punteggi per avere una precisa idea del significato dei singoli punteggi ottenuti.

Il vantaggio di ciò, ovvero quello di paragonare punteggi (di persone diverse, o della stessa persona in due momenti diversi) è quello che non serve effettuare calcoli ma basta utilizzare i dati grezzi.

A questo punto il processo è concluso e l’obiettivo di ottenere una scala di misura valida ed affidabile è stato raggiunto; un processo che, attuato nella sua interezza, è certamente articolato e

costoso per il numero ed anche le complessità delle sue fasi, ma la qualità della ricerca dipende strettamente dalla precisa esecuzione del percorso appena descritto.

Il lavoro di Churchill e Gilbert rappresenta dunque una pietra miliare per quanto riguarda lo sviluppo di scale di misura: introduce i concetti imprescindibili di affidabilità e validità, fornisce in generale un ben strutturato e giustificato percorso da seguire per sviluppare una buona misura; è insomma un punto di riferimento per tutti quelli che in seguito si troveranno ad affrontare lo stesso argomento, per cercare di apportare dei miglioramenti o per aggiungere nuovi elementi o nuove teorie.

1.4 Supporto alla teoria di Churchill: lo studio di Peter sull'affidabilità

Nello stesso numero del *Journal of Marketing* nel quale Churchill descrive la prima teoria strutturata sul processo di sviluppo di scale di misura (1979) compare un non meno importante articolo ad opera di J. P. Peter: “*Reliability: A Review of Psychometric Basics and Recent Marketing Practices*”.

Si tratta di uno studio che si focalizza fundamentalmente sulla proprietà di affidabilità che una misura deve possedere partendo dalla constatazione che nel passato poca era la cura dei ricercatori riguardo questa proprietà e di conseguenza mancava anche in questo caso un solido impianto teorico.

Il lavoro di Peter integra dunque quello di Churchill per formare assieme una teoria che sarà di riferimento negli anni successivi per i ricercatori di marketing, ponendo l'attenzione per la prima volta sull'affidabilità, condizione necessaria anche se non sufficiente per l'accertamento dell'altra proprietà fondamentale di validità e in generale per la realizzazione di una buona scala di misura.

1.4.1 Approcci tradizionali all'affidabilità

L'approccio di base più quotato sino ad allora era quello che si riferiva a quanto teorizzato da Spearman agli inizi del novecento.

Secondo questa teoria, la media e la varianza di un livello di punteggio ottenuto dalla somministrazione della scala (formata da più item) sono costituiti dalla somma di due componenti, una che riflette il vero punteggio ed un'altra che consiste in un termine di errore che determina quindi una discrepanza tra la corretta misurazione del fenomeno e l'effettiva misurazione attraverso la scala. La grandezza dell'errore determina il grado di affidabilità di una misura, ed è

causato fondamentalmente dalla presenza di item che non appartengono propriamente al dominio della variabile di interesse. Da ciò si perviene ad un primo indice che misura l'affidabilità: non si fa altro che considerare il rapporto tra varianza osservata e varianza vera nell'impossibilità di avere quest'ultima e, sfruttando la sua scomposizione in due componenti, si perviene a definire questo indice come uno meno il rapporto tra la varianza dell'errore e la varianza osservata.

L'idea di Spearman di stabilire quanta parte della varianza osservata è sistematica è alla base di tre metodi per accertare l'affidabilità che in seguito vennero sviluppati: test-retest, consistenza interna, altre forme alternative. Questi tre approcci hanno in comune l'intento di valutare la correlazione del punteggio di una scala con quello di altre scale diverse a seconda di quale metodo si adotti.

Per quanto riguarda il test-retest il punteggio con cui viene correlato il punteggio di partenza è quello che risulta dalla somministrazione della stessa scala agli stessi soggetti alle stesse condizioni dopo un certo intervallo di tempo da quella iniziale (solitamente due settimane). Questo metodo, sebbene utile, presenta alcune problematiche: generalmente l'affidabilità tende a diminuire all'aumentare della distanza di tempo, inoltre il fenomeno osservato potrebbe subire delle modifiche durante l'intervallo di tempo, e da ultimo c'è da considerare la presenza nella correlazione di una parte dovuta alla correlazione tra item identici.

Il test-retest può allora essere integrato con la verifica della coerenza interna, durante la quale la correlazione di interesse è quella tra item di una stessa scala somministrata in un dato istante. Questa procedura prevede la divisione degli item in due gruppi casuali, i cui punteggi vengono confrontati attraverso il calcolo della loro correlazione. Il problema ovvio con cui ci si scontra è quello relativo alla scelta dei due gruppi di item: a seconda della loro composizione si possono infatti ottenere risultati diversi. Questo ostacolo può essere superato calcolando un indice che tiene conto di tutte le divisioni di item che possono essere effettuate e che viene molto usato: l'alfa di Cronbach.

L'ultimo metodo tradizionalmente usato per l'accertamento dell'affidabilità propone di utilizzare due scale simili da somministrare agli stessi soggetti in due momenti diversi. Le due scale devono ovviamente avere delle caratteristiche comuni ma, allo stesso tempo, devono presentare delle piccole differenze. In questo caso i problemi riguardano prima di tutto il modo stesso di trovare una misura che sia sostanzialmente analoga alla prima; in secondo luogo, bisogna provare l'equivalenza dei contenuti delle due scale, visto che se, ad esempio, trovo una bassa correlazione tra le due, non si può stabilire se esse abbiano un'intrinseca bassa affidabilità o se una delle due non presenta similarità di contenuti con l'altra.

1.4.2 Creazione di una teoria generalizzata

A questo punto, dopo aver passato in rassegna i vari approcci usati in passato per la verifica dell'affidabilità, Peter getta le basi per la costruzione di una teoria generalizzata.

Peter parte dalla premessa che i metodi tradizionali non tengono conto del fatto che le fonti di errore che determinano l'inaffidabilità possono essere molteplici: esse solo in parte vengono individuate.

Per questo motivo alla base della teoria di Peter c'è il concetto di campionamento di quelle che chiama condizioni di misura: sono un aspetto del processo della misura, come lo sono ad esempio gli item usati nella scala o la misura stessa o ancora i tempi e i luoghi di misura e, attraverso il loro campionamento, ci si chiede se esse possano rappresentare tutte le condizioni possibili che possono verificarsi.

Non tutte le condizioni possono essere contemporaneamente colte per questo, attraverso un loro campionamento casuale, viene garantita l'attendibilità dei vari indici che vengono calcolati (correlazione, media, varianza) in un'ottica di conseguimento di risultati generalizzabili.

In questo modo un metodo tradizionale come il test-retest deve avere a che fare con la capacità di poter generalizzare i punteggi ottenuti ad altri universi di ricerca: un esito positivo del test infatti (ovvero alta correlazione tra score) non assicura l'efficacia della misura in altri contesti.

Per rendere possibile qualsiasi sorta di generalizzazione è dunque necessario implementare nuove procedure di misura, tenendo comunque conto che il principale obiettivo resta quello di indagare su ciò che causa la variabilità osservata, in particolare di cercare di scoprire e cogliere tutti quei fattori che stanno alla base della varianza d'errore che genera la discrepanza tra il vero punteggio e il punteggio osservato.

Peter propone allora di espandere il modello dell'analisi della varianza per includere le diverse fonti che causano la variabilità osservata: la varianza totale è quindi vista come l'insieme di componenti quali il tempo (il momento in cui viene effettuata la somministrazione), lo strumento, e tutti quegli aspetti che fanno parte del concetto di "condizione di misura" e che secondo il ricercatore possono essere fonte di errore.

L'importanza di questa innovazione sta nel fatto che si è in grado di stabilire il peso dei singoli aspetti nel determinare la varianza totale osservata: ciò infatti non è conseguibile attraverso i metodi tradizionali i quali, per la loro natura, non considerano in partenza tutti quei fattori che soggiacciono al gap tra variabilità osservata e variabilità effettiva di un fenomeno (il coefficiente alfa, ad esempio, considera unicamente il fattore item).

Al di là dell'utilità dell'analisi della varianza "estesa" bisogna comunque valutare caso per caso il suo impiego: per prima cosa infatti, all'aumentare delle fonti di errore prese in considerazione

aumenta anche il grado di complessità dell'interpretazione e della costruzione del modello; in secondo luogo, il ricercatore può avere dei validi motivi per escludere certi fattori già in partenza, perché effettivamente ed oggettivamente ininfluenti.

1.4.3 Affidabilità in precedenti studi di marketing

Come ultimo passo, Peter esegue una rassegna dei passati studi condotti da altri ricercatori di marketing sotto la lente di ingrandimento fornita dalla sua teoria: lo scopo di questa operazione non è quello di criticare i risultati ottenuti e la qualità delle indagini, quanto piuttosto di portare alla luce alcune carenze ed alcuni problemi.

Sono stati selezionati alcuni studi appartenenti ad aree diverse pubblicati sulle più importanti riviste di marketing e si è puntata l'attenzione sui tipi di scala utilizzati, sui metodi per accertare l'affidabilità, sul numero degli item e sulla numerosità campionaria degli intervistati.

Dall'analisi di queste variabili sono dunque emerse diverse problematiche potenzialmente risolvibili se si adotta l'approccio della teoria di Peter: alcune di carattere generale, quale il metodo per trovare il numero di item giusti e la numerosità campionaria opportuna o il livello di un indice dopo il quale stabilire l'affidabilità di una scala; altre problematiche invece di carattere particolare che affliggono i singoli studi, quali il breve periodo fatto passare tra test e retest, la differenza di composizione del campione rispondente tra il test e il retest, oppure l'uso del metodo split-half per il calcolo della coerenza interna, i cui svantaggi sono stati descritti in precedenza.

Il lavoro di Peter, che si basa prevalentemente su una teoria che mira ad analizzare le diverse fonti di errore tramite una analisi della varianza "estesa", è importante per aver portato alla ribalta il problema dell'accertamento dell'affidabilità fornendo degli strumenti per ottenerla: fino ad allora infatti poca era stata la cura per la verifica di una proprietà che tuttavia è imprescindibile per l'ottenimento di una buona misura.

Peter al pari di Churchill ha quindi il merito di aver dato un notevole contributo alla teoria dello sviluppo di scale di misura valide ed affidabili, facendo fare un sensibile salto di qualità a studi teorici e ricerche di marketing che verranno effettuati negli anni successivi.

Capitolo 2: Sviluppo ed evoluzione delle teorie

2.1 Introduzione

Il paradigma di Churchill ha rivoluzionato il modo di sviluppare le scale di misura a tal punto che esso è stato un costante punto di riferimento per tutti gli studi successivi che hanno trattato la misura di qualche variabile di marketing.

Per una vera e propria messa in discussione del modello churchilliano si deve così aspettare il 2002, anno in cui Rossiter propone un nuovo modo di misurare i costrutti di marketing nell'articolo *"The C-OAR-SE procedure for scale development in marketing"* apparso sul *Journal of Research in Marketing*.

La procedura di Rossiter è poi il punto di partenza per un'analisi storica di quanto è stato prodotto in termini di scale di misura e una conseguente esposizione di un nuovo approccio: *"How fine is C-OAR-SE? A Generalizability Theory perspective on Rossiter's procedure"* (2005) A. Finn e U. Kayande propone la teoria di generalizzabilità multivariata.

Precedentemente al lavoro di Rossiter sono da segnalare altri due importanti studi: *"An Update Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment"* (1988) di David W. Gerbing e James C. Anderson, che introduce gli strumenti per studiare le variabili unidimensionali e *"Index Construction with Formative Indicators: an alternative to Scale Development"* (2001) di A. Diamantopoulos e M. Winklhofer, che introduce un metodo alternativo di misura attraverso gli indicatori formativi.

2.2 Un aggiornamento del paradigma churchilliano: lo studio di variabili unidimensionali di Anderson

Un aggiornato modello di sviluppo di scale di misura viene presentato nel 1988 da David W. Gerbing e James C. Anderson sul *Journal of Marketing Research* con l'articolo *"An Update Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment"*.

Questo lavoro, che si basa sulle teorie di Churchill e Gilbert (1979), parte dal presupposto dell'impossibilità di misurare una variabile di marketing con una misura a singolo item. Da ciò la

necessità di usare più indicatori diversi in grado di cogliere ciascuno una e una sola dimensione, per ottenere così una misura del costrutto attraverso la somma dei vari risultati ottenuti. L'unidimensionalità degli indicatori è necessaria affinché la proprietà di validità possa essere soddisfatta.

Per accertare l'unidimensionalità lo strumento che viene ritenuto fondamentale è quello dell'analisi fattoriale confermativa, considerato un metodo molto più efficace rispetto ad altri quali ad esempio il coefficiente alfa, l'analisi fattoriale esplorativa e la correlazione item-totale.

2.2.1 Gli elementi distintivi della teoria: unidimensionalità e analisi fattoriale confermativa

Gli elementi distintivi che stanno alla base della teoria presentata in questo articolo sono i concetti di misura unidimensionale e di analisi fattoriale confermativa.

- *Misure unidimensionali*

L'idea cardine su cui poggia il modello di sviluppo di scale di Gerbing ed Anderson è quello di misura unidimensionale.

Per spiegare questa nozione viene introdotta la seguente relazione lineare:

$$x_i = a_i * k + e_i ,$$

dove x_i è la i -esima misura o indicatore, k il singolo fattore considerato, a_i il peso corrispondente del fattore ed e_i il corrispondente residuo della regressione.

Da questa relazione si derivano i seguenti due criteri, che rappresentano una condizione necessaria per l'unidimensionalità:

$$\rho_{ij} = \rho_{ik} * \rho_{jk}$$

e

$$\rho_{ii^*} = \rho_{ik} * \rho_{kk^*} * \rho_{i^*k}$$

Il primo rappresenta la proprietà di coerenza interna e riguarda la correlazione tra due strumenti che misurano la stessa caratteristica (ρ_{ij}).

Il secondo invece descrive la proprietà di coerenza esterna, e presenta la correlazione tra due strumenti che misurano due diverse variabili (ρ_{ii^*}).

- *Analisi fattoriale confermativa*

L'analisi fattoriale confermativa di un modello di misura ad item multipli è il vero elemento innovativo introdotto da Gerbing ed Anderson nello sviluppo di scale di misura, e rappresenta il fattore distintivo della teoria qui esposta.

Le precedenti procedure, che si basavano sul coefficiente alfa, sull'analisi fattoriale esplorativa, e sulla correlazione item-totale, non vengono qui ritenute in grado di accertare l'unidimensionalità al pari della tecnica dell'analisi fattoriale confermativa.

La procedura proposta consiste quindi nella stima dei parametri del modello sopra presentato nella descrizione dell'unidimensionalità e nel calcolo delle correlazioni attraverso i due criteri di coerenza interna ed esterna ed, infine, nell'analizzare i dati ottenuti traendo le opportune conclusioni.

Una semplice regola da osservare in questa analisi è, ad esempio, quella di eliminare gli indicatori che presentano una correlazione significativamente diversa da zero con un altro fattore diverso da quello che si sta studiando: è ovvio che se si vuole cogliere l'aspetto unidimensionale non bisogna ammettere qualsiasi indicatore (item) che abbia qualche legame con altri fattori.

Alla fine di questa analisi si otterrà un insieme di item capace di cogliere efficacemente l'unidimensionalità della variabile sulla quale si sta indagando, grazie all'accertamento dei criteri di coerenza interna ed esterna sopra descritti.

2.2.2 Confronto tra analisi fattoriale confermativa e altri metodi

Vengono ora brevemente descritti tre tipi di analisi che si effettuano nell'ambito dello sviluppo di scale, ovvero lo studio della correlazione item-totale, l'analisi fattoriale esplorativa e il coefficiente alfa, e viene messa in luce la loro scarsa efficacia a cogliere l'unidimensionalità di una variabile rispetto all'analisi fattoriale confermativa.

- *La correlazione item-totale*

Nel passato diversi studiosi hanno riconosciuto l'importanza della correlazione tra il punteggio di un singolo item con il punteggio totale, preferendola ad altri tipi di analisi.

Tuttavia, e qui sta la differenza con l'analisi fattoriale confermativa, Gerbing e Anderson fanno notare che questi tipi di analisi non possono dire nulla sulla coerenza esterna, non facendo risaltare ad esempio la differenza tra insiemi di indicatori che stimano fattori diversi ma correlati.

- *L'analisi fattoriale esplorativa*

La tecnica dell'analisi fattoriale esplorativa è usata per selezionare un appropriato insieme di item (o indicatori) che individuino le dimensioni di una variabile o, nell'ambito unidimensionale, il singolo fattore.

Essa viene implementata nella fase iniziale di uno studio con lo scopo di ottenere una prima rappresentazione generale del costrutto di interesse nelle sue componenti. Le differenze con l'analisi fattoriale confermativa sono essenzialmente due: da un lato mancano per l'esplorativa delle relazioni matematiche a supporto, dall'altro lato esse sostanzialmente avvengono in due fasi diverse della procedura con scopi diversi, essendo l'esplorativa una sorta di indagine preliminare e la confermativa un mezzo per stabilire in via definitiva la natura della variabile.

- *Il coefficiente alfa*

Una volta conseguita l'unidimensionalità non è detto che si abbia ottenuto una scala affidabile.

Per accertare l'affidabilità entra in gioco il coefficiente alfa, il quale tuttavia spesso è stato usato come indice per verificare l'unidimensionalità.

Ma unidimensionalità ed affidabilità sono due concetti distinti: il primo ha a che fare con la correlazione tra indicatori, mentre il secondo ha a che fare con il numero di item e con l'affidabilità degli stessi item. In più, quando si va a calcolare il coefficiente alfa, si dà per assunto tra le altre cose l'unidimensionalità che caratterizza l'insieme di item.

L'obiettivo che Gerbing e Anderson danno al loro studio è fondamentalmente quello di fornire degli strumenti teorici per studiare l'unidimensionalità di una scala; è da ricordare d'altro canto che ciò va ad inserirsi in un contesto più ampio di sviluppo di una scala che non può prescindere dalla verifica della proprietà di validità del costrutto.

Precisato ciò, alla luce di quanto analizzato nella descrizione della ricerca, lo strumento per eccellenza è qui costituito dall'analisi fattoriale confermativa: esso rappresenta infine il miglior mezzo di cui ci si può servire se il nostro obiettivo è quello di misurare un costrutto caratterizzato dall'unidimensionalità, e va quindi ad aggiornare l'originario paradigma fornito da Churchill e Gilbert nel 1979.

2.3 Un'alternativa alle scale di misura: gli indicatori formativi

La letteratura esistente individua un'altra via per la misura di un costrutto di marketing che cerca di sganciarsi dagli studi eseguiti da Churchill e Peter (1979).

Questa nuova teoria è esposta nell'articolo "*Index Construction with Formative Indicators: an alternative to Scale Development*" (2001) ad opera di A. Diamantopoulos e M. Winklhofer apparso sul *Journal of Marketing Research*.

L'approccio alternativo che supera la metodologia classica ha come concetto cardine quello di indicatore formativo e come strumento da sviluppare un semplice indice evidentemente in opposizione con le "vecchie" scale.

Diamantopoulos e Winklhofer ribaltano lo schema in base al quale i singoli item di una scala rappresentano una serie di effetti o manifestazioni di una variabile soggiacente il costrutto in analisi, proponendo al contrario degli indicatori che sono appunto formativi nel senso che costituiscono essi stessi delle variabili che causano ed insieme definiscono un variabile latente.

L'intento dei due autori è dunque quello di fornire un diverso modo di affrontare il problema della misura per quelle situazioni in cui il metodo tradizionale si rivela poco efficace e si vuole inoltre fornire una base teorica completa cercando di sintetizzare e sviluppare quanto già presente in altri studi che hanno trattato la stessa materia.

2.3.1 Un approccio teorico

La necessità di costruire questi particolari indicatori sta nel fatto che spesso la variabile latente che si vuole analizzare è causata da una serie di elementi diversi tra loro: in questi casi un approccio tradizionale, basato su una concezione opposta (ovvero la stessa variabile latente passa ad essere una causa), si dimostrerebbe inutile. In sostanza ci sono delle variabili che sono la manifestazione di un gruppo di componenti che sommate contribuiscono a definire la natura e il carattere della variabile stessa, che diventa quindi il risultato di una serie di aspetti che allo stesso tempo restano comunque autonomi e indipendenti tra loro.

Un esempio è ben rappresentato dallo stato socioeconomico di una persona che per sua natura dipende ed è quindi effetto di una serie di cause quali ad esempio il reddito o l'istruzione; se uno di questi aspetti dovesse variare subirebbe un cambiamento anche lo stato socioeconomico, ma l'incremento di quest'ultimo non sarebbe univocamente individuabile come un aumento di una specifica variabile. In questo caso, come in molti altri in ambito economico, l'approccio formativo costituisce sicuramente la scelta migliore.

Di contro un costrutto come il coinvolgimento non può essere considerato al pari dello stato socioeconomico: la scala che si va a costruire per coglierlo infatti è costituita da item che hanno lo scopo di riflettere il costrutto analizzato e non di spiegarlo ponendosi come una delle diverse cause come avviene per gli indicatori formativi, anche perché le cause possono essere certamente riconducibili a diversi ambiti ma non possono essere altrettanto analiticamente definite ed individuate a tal punto da creare separati ed autonomi indicatori.

Sotto questa prospettiva dunque la variabile latente coincide con la sua misura, o al massimo ne è una sua funzione, dove ciascuna delle componenti-cause ha un certo peso nel determinarla. La seguente relazione spiega il concetto descritto:

$$(1) \quad A = b_1x_1 + b_2x_2 + \dots + b_nx_n + e \quad ,$$

dove A rappresenta la variabile latente, le x_i variabili osservate con i loro rispettivi pesi b_i , e termine d'errore.

L'approccio formativo degli indicatori si distingue sotto molti aspetti da quello delle classiche scale di misura. La prima fondamentale è che per le misure formative è impossibile scambiare un indicatore con un altro perché si perderebbero fondamentali informazioni sul costrutto, quando invece per le misure riflesse la sostituzione di un item con un altro non provoca l'allontanamento dall'obiettivo di cogliere la variabile (tant'è che a volte la sostituzione è di prassi per ottenere un coefficiente alfa buono). In secondo luogo, l'approccio formativo non permette di trovare la correlazione tra gli indicatori, essendo questi variabili esogene, creando dei problemi nell'accertamento della validità degli stessi. Terzo, il problema della coerenza interna: se due indicatori sono ad esempio correlati negativamente non si deve per forza concludere che essi non siano in grado di descrivere il costrutto in analisi. Quarto, il termine d'errore è solo di disturbo e non presenta correlazione con gli indicatori formativi quando invece nelle misure riflesse la correlazione è presente. Da ultimo, il modello (1) non è stimabile se non aggiungendo altre variabili che sono causate dal costrutto in questione; con l'altro approccio invece non servono aggiunte ma solo una analisi fattoriale confermativa per stimare i parametri.

La conseguenza di tutto ciò, come facilmente intuibile, è che i metodi tradizionali di accertamento delle proprietà di affidabilità e validità possono rivelarsi inefficaci, presentandosi dunque la necessità di trovare nuove strade: risultano infatti inutili la coerenza interna già citata per l'affidabilità e la validità convergente e discriminante per la validità della misura.

2.3.2 La costruzione dell'indice

Il processo di sviluppo di misure riflesse da Churchill in poi è sempre stato costituito da una sequenza precisa di stadi da attraversare per raggiungere una scala definitiva. Per quanto riguarda invece gli indicatori formativi, l'individuazione di un percorso da seguire non è facile, ma visionando i diversi studi sull'argomento Diamantopoulos e Winklhofer riescono a comporre un percorso a quattro livelli: specificazione del contenuto, specificazione dell'indicatore, collinearità dell'indicatore, validità esterna.

La prima fase di specificazione del contenuto, comune al metodo tradizionale, è fondamentale e molto delicata visto che tutte le cause della variabile latente devono essere individuate per evitare di ottenere uno strumento che sarebbe scarsamente informativo.

Nella specificazione degli indicatori non è ovviamente applicabile il modello del campionamento degli item utilizzato nelle tradizionali scale: ancora una volta si vuole sottolineare l'insostituibilità delle singole componenti, senza tuttavia escludere una purificazione della misura che deve essere comunque molto prudente.

L'attenzione posta sulla collinearità tra gli indicatori è dovuta al fatto che se essa è presente in maniera forte non è più possibile distinguere correttamente i vari effetti delle componenti sulla variabile latente. Ciò che si può fare è invece escludere quegli indicatori che risultano combinazione lineari di altri per evitare ridondanze. D'altra parte anche il valore dei pesi degli indicatori può essere usato per accertare la validità degli indicatori stessi: un coefficiente alto ad esempio testimonia il forte legame con la variabile latente.

Da ultimo, tenuto conto dei problemi con cui ci si scontra nell'accertamento delle proprietà che una misura deve avere, si cerca di verificare la validità esterna per stabilire quali e quanti indicatori si debbano tenere per realizzare la misura. A questo scopo si possono adottare diversi approcci: si può creare un item riassuntivo dei contenuti degli indicatori e si valuta poi la sua correlazione con i singoli indicatori; possono essere inserite nel modello alcune variabili endogene per far sì che il modello sia identificabile stabilendo poi la bontà del modello stesso; si può infine confrontare la variabile latente con altri costrutti con i quali essa ha un legame accertato.

Il lavoro di Diamantopoulos e Winklhofer non ha solo il merito di aver creato una sintesi tra tutti gli studi che in maniera meno sistematica avevano discusso l'argomento, ma ha anche e soprattutto quello di aver fornito uno strumento che va a sopperire ad un vuoto che per natura è lasciato dal metodo tradizionale, quello che riguarda quelle variabili latenti che sono causate da una serie di altre variabili endogene.

Sta dunque al ricercatore stabilire la natura del costrutto col quale si trova a lavorare per decidere se preferire l'approccio riflessivo o l'approccio formativo; è ovvio quanto necessario quindi un forte sforzo a livello di teorizzazione di costrutto.

In ultima analisi, e qui sta la vera importanza dello studio di Diamantopoulos e Winklhofer, è auspicabile il trattamento di alcune variabili di marketing (campo "dominato" da misure riflessive) con un approccio formativo per ottenere non solo modelli che migliorano l'adattamento, ma anche misura decisamente e sostanzialmente migliori.

2.4 La procedura C-OAR-SE

Una nuova procedura per lo sviluppo di scale di misura viene presentata da Jhon R. Rossiter sull'*International Journal of Research in Marketing* con l'articolo "*The C-OAR-SE procedure for scale development in marketing*" (2002).

Il bisogno di una nuova teoria nasce dalla constatazione di alcuni limiti del modello tradizionale di sviluppo di scale fino ad allora seguito, ovvero quello di Churchill datato 1979: eliminazione di item fondamentali nella concettualizzazione dell'oggetto di studio per perseguire l'unidimensionalità, l'aggiunta inopportuna di item per avere un coefficiente alfa elevato, e il solo uso di alfa elevati per provare la validità della scala.

La procedura C-OAR-SE si pone dunque l'obiettivo di superare questi limiti adottando per prima cosa un approccio razionale aprioristico: l'attenzione viene infatti posta sulla validità dei contenuti attraverso una concettualizzazione del costrutto che avviene nei primissimi stadi del processo di sviluppo della scala e che tende ad essere indipendente da eventuali riscontri empirici ottenuti con la somministrazione della scala stessa.

2.4.1 L'impianto teorico della procedura C-OAR-SE

Il nome C-OAR-SE è un acronimo e sta ad indicare le varie fasi del processo di sviluppo di una scala di misura: la definizione del Costrutto, la classificazione dell'Oggetto, la classificazione dell'Attributo, l'identificazione del Valutatore, la formazione della Scala, Classificazione e Report.

- *Definizione del costrutto*

La teoria C-OAR-SE prevede una precisa e rigorosa definizione del costrutto; d'altra parte l'importanza di questa fase era già stata sottolineata da Churchill (1979).

Il costrutto viene dunque descritto in tre parti principali: l'oggetto e le sue componenti, l'attributo e le sue componenti e l'unità valutatrice. Ne è un esempio "La qualità dei servizi di IBM percepita dai clienti", dove l'oggetto è "IBM", l'attributo "la qualità dei servizi" e l'unità valutatrice "i clienti".

Per quanto riguarda questa fase della procedura Rossiter attinge a precedenti lavori di altri studiosi.

- *Classificazione dell'oggetto*

La classificazione dell'oggetto coinvolge un gruppo di esperti assistiti da delle interviste con i soggetti valutatori, per individuare una delle seguenti definizioni dell'oggetto del costrutto:

-singolare concreto: quando i valutatori conoscono l'oggetto, il quale è singolare. È il caso di "qualità del servizio di IBM" dove l'oggetto, IBM, è unico e descritto e conosciuto allo stesso modo da tutti i valutatori. In questa definizione rientrano insieme di oggetti omogenei, come può essere la Coca-Cola (che viene venduta in diversi formati).

-collettivo astratto: quando l'oggetto viene percepito come un insieme di entità eterogenee . A questa categoria appartiene ad esempio "soft drinks", un insieme astratto che contempla elementi eterogenei tra loro.

-astratto composto: quando i valutatori danno interpretazioni dell'oggetto diverse tra loro in quanto riconoscono in esso diverse componenti. L'oggetto "Capitalismo" può rientrare in questa categoria: il giudizio su di esso può variare a seconda di quale sua componente si consideri (libertà di impresa,ricerca di profitti,...).

- *Classificazione dell'attributo*

Il terzo passo della procedura (che è anche il più complesso) consiste nella classificazione dell'attributo, ovvero di ciò su cui il valutatore esprime il proprio giudizio.

Sono tre le categorie in cui esso si distingue:

-concreto: come nel caso dell'oggetto, si parla di attributo concreto quando c'è unanimità da parte dei valutatori sul modo di intendere l'attributo stesso (ad esempio "propensione all'acquisto", "qualità di un prodotto",...). Per questi tipi di attributi basta un singolo item.

-composto: quando più componenti formano l'attributo ("qualità del servizio",dove il servizio è visto come un insieme di elementi). Si dovrà usare quindi un numero di item adeguato per cogliere le diverse componenti.

-astratto: quando l'attributo è astratto e gli item ad esso legati rappresentano una manifestazione indicativa della disposizione o stato del valutatore, ovvero attraverso la somministrazione degli item si cerca di estrarre l'attributo. Ne sono un esempio "l'orientamento del cliente" o "il coinvolgimento personale".

- *Identificazione del valutatore*

Scegliendo la prospettiva dalla quale si esprime un giudizio, ovvero identificando l'entità valutatrice, si conclude la fase di definizione del costrutto. I tipi di valutatore sono tre:

-individuale: associato ad esempio a costrutti come il coinvolgimento o l'attitudine personale.

-gruppo: insieme di consumatori, venditori, compratori, fornitori, ecc.

-esperti: fondamentalmente dei giudici, il cui grado di accordo viene correlato con l'affidabilità.

- *Formazione della scala*

A questo punto la procedura prevede la formazione della scala; si tratta di mettere assieme le parti degli item riguardanti l'oggetto con le parti riguardanti l'attributo, e di creare i cosiddetti "stems" e "leaves" che compongono il singolo item, ovvero rispettivamente la domanda e le alternative di risposta.

Il processo C-OAR-SE abbraccia la visione "single-item", in netto contrasto con la "multiple-item" caratteristica di Churchill: Rossiter spiega infatti che basta un item per cogliere un costrutto di marketing (più item solo se si tratta di coprire le varie dimensioni di un oggetto/attributo) e che la scelta di Churchill era dovuta al fatto che egli mirava fondamentalmente a costruire dei test di abilità, ovvero proponeva item di diversi livelli di difficoltà che si scontravano con la capacità di rispondere da parte dell'intervistato.

Il passo successivo è quello di fare un pre-testing attraverso un'intervista cognitiva, per accertare che gli item abbiano il significato che il ricercatore si era proposto di dare loro. In questa fase una particolare attenzione deve essere poi dedicata ad i nuovi attributi "estratti" che si vanno a concettualizzare: in questo caso è necessario infatti che degli esperti giudichino il prototipo di scala prima che questa venga definitivamente applicata.

Un altro aspetto fondamentale della formazione della scala riguarda il formato della risposta. La scala di Likert ad esempio è sconsigliabile, in quanto l'intensità è posta nella domanda e ciò può creare delle ambiguità. L'item ottimale è dunque composto da una domanda che non includa

l'intensità e da una risposta in cui l'intensità venga categorizzata tra un minimo ed un massimo (con delle probabilità, delle frequenze o dei gradi).

L'ultimo passo nella formazione della scala è quello di fare in modo che gli item compaiano in ordine casuale: un'azione necessaria al fine di rendere minima la correlazione tra risposte e di ottenere quindi un risultato non "artefatto".

- *Classificazione*

Dopo aver somministrato la scala è necessario stabilire quali regole seguire per calcolare il punteggio totale.

Un primo aspetto da sottolineare è che le regole possono cambiare a seconda della natura dell'oggetto e dell'attributo.

In generale si può, ad esempio, sommare semplicemente i singoli punteggi di ciascun item, stabilire un livello minimo da raggiungere per singolo item o calcolare una media ponderata, dando pesi diversi ai vari item.

- *Report*

Nella fase finale di reporting, in cui vengono sintetizzati i risultati dell'indagine, un accorgimento utile da adottare è relativizzare ad esempio in una scala da uno a dieci i punteggi assoluti ottenuti dalla somministrazione della scala, prendendo come riferimenti il valore più basso e quello più alto.

2.4.2 Accertamento delle proprietà di validità ed affidabilità

- *Validità*

La validità del costrutto è un concetto che sta alla base della teoria C-OAR-SE, in quanto è ciò su cui si punta per dimostrare che la procedura qui descritta produce scale di misura migliori alle precedenti. Due in particolare sono i tipi di validità su cui ci si sofferma: validità del costrutto e validità attesa.

Per verificare la validità del costrutto Churchill proponeva la cosiddetta "multitrait-multimethod matrix", che in sostanza rappresentava le correlazioni tra le diverse componenti di una variabile misurate con metodi diversi.

Tuttavia, nel caso della procedura C-OAR-SE, non può essere usato il metodo sopra descritto: prima di tutto è necessaria la certezza del fatto che la misura vecchia sia essa stessa valida in partenza e, in secondo luogo, alte o basse correlazioni non possono dire quale delle due misure sia migliore.

La “multitrait-multimethod matrix” non fornisce prove risolutive per accertare la validità del costrutto: per raggiungere questo obiettivo si deve invece fare affidamento sulla validità dei contenuti, ovvero ci si deve basare sul lavoro svolto all’inizio nella definizione del costrutto.

La validità attesa non è qui interpretata come tentativo di massimizzare la correlazione tra la procedura C-OAR-SE e un qualche criterio scelto, attraverso l’inserimento di adeguati item.

L’orientamento è piuttosto quello di stabilire la validità della scala indipendentemente dal costrutto, seguendo semplicemente la procedura C-OAR-SE.

- *Affidabilità*

Come descritto in passato da Churchill (1979), l’affidabilità è propria di una scala se sono minimi gli errori casuali derivati ad esempio da fattori ambientali o personali legati al soggetto valutatore.

Rossiter spiega che il punto di partenza nell’accertare l’affidabilità resta la validità dei contenuti, concetto la cui importanza è già stata ribadita più volte in precedenza. In più egli afferma che l’affidabilità è da considerarsi in relazione al punteggio più che alla scala e, cosa non meno importante, il soggetto valutatore influisce sull’affidabilità stessa.

Tenuto conto che il coefficiente alfa di Cronbach non è sufficiente e che il test-retest può non essere efficace, il problema dell’affidabilità va allora affrontato distinguendo tra i seguenti tipi di soggetti valutatori:

-valutatori individuali

Nel caso di attributi concreti e composti, non ci si pone il problema dell’affidabilità: il valutatore dà infatti un punteggio preciso, ammettendo che egli non dia volontariamente risposte false né che sia incompetente. Nel caso invece di attributi astratti, data la loro particolare natura, si deve far riferimento al coefficiente alfa.

-valutatori esperti

Nel caso di valutatori esperti, la precisione del punteggio (e quindi l’affidabilità) deriva dal grado in cui gli esperti stessi sono in accordo tra di essi. Il numero di esperti coinvolti o il numero di categorie degli attributi possono ad esempio influenzare la valutazione finale.

-gruppo di valutatori

Quando l'entità valutatrice è costituita da un gruppo, la numerosità di quest'ultimo è strettamente legata all'affidabilità del punteggio: più è grande il campione e maggiore sarà la precisione del punteggio (e più ampi gli intervalli di confidenza andranno a considerare).

Con la trattazione del problema dell'affidabilità si conclude la descrizione della procedura di sviluppo di scale di misurazione di Rossiter. Una procedura innovativa rispetto a quanto è stato proposto in passato (temporalmente essa compare 23 anni dopo il primo importante lavoro sull'argomento, quello di Churchill del 1979), che basa il suo successo su un approccio razionale più che empirico, radicandosi fortemente sull'iniziale definizione del costrutto che si vuole studiare: è la validità dei contenuti quindi, verificata e studiata per prima, ciò su cui si punta per dimostrare la bontà della procedura C-OAR-SE e la sua migliore capacità di cogliere la variabile di interesse.

2.5 La teoria di generalizzabilità multivariata

"How fine is C-OAR-SE? A Generalizability Theory perspective on Rossiter's procedure" (2005) appare sul *Journal of Research in Marketing* ad opera di A. Finn e U. Kayande e rappresenta una summa di quanto la letteratura sulle procedure di sviluppo di scale aveva prodotto sino a quel momento.

In questo articolo infatti viene fatta una panoramica sui modi in cui il problema della misura di costrutti è stato affrontato prendendo come ovvi punti di riferimento la teoria di Churchill del 1979 e quella più recente di Rossiter del 2002.

In questo studio, oltre a questo impegno descrittivo che ha lo scopo di far emergere vantaggi e svantaggi delle differenti procedure, Finn e Kayande si pongono l'obiettivo di creare a loro volta una nuova procedura che in qualche modo da un lato riesca a superare e dall'altro riesca a sintetizzare rispettivamente i difetti e i pregi delle vecchie scale.

L'approccio che viene usato nel perseguire la costruzione della scala è quello della teoria di generalizzabilità multivariata attraverso il quale si va dunque a porre rimedio a quei problemi che le altre procedure hanno e che sono legati alla loro natura e struttura.

2.5.1 Rassegna storica delle procedure di sviluppo di scale di misura

Il modello introdotto da Churchill ha rappresentato senza dubbio negli anni successivi alla sua pubblicazione un impareggiabile punto di riferimento per chi nel campo del marketing si è di volta in volta trovato ad affrontare problemi di misura di variabili.

Prima di allora mai nessuno aveva elaborato una teoria così solida e strutturata e le scale che si costruivano erano spesso utili a singole peculiari situazioni e, cosa più importante, non venivano poste particolari attenzioni ad aspetti delicati quali ad esempio la verifica delle fondamentali proprietà di affidabilità e validità.

Anche se l'avvento della teoria di Churchill ha fatto evidentemente compiere un significativo salto di qualità agli studi sulla realizzazione di scale, allo stesso tempo si evidenzia comunque un'importante problematica: Finn e Kayande affermano infatti che l'approccio prettamente empirico di Churchill, mirato principalmente al conseguimento dei "numeri giusti", ha lasciato poco spazio all'importante aspetto della concettualizzazione del costrutto.

Ecco quindi che la procedura ideata da Rossiter nel 2002 va a colmare questa importante lacuna attraverso una definizione del costrutto molto specifica e dettagliata che contempla l'individuazione di un oggetto, un attributo ed un valutatore.

Tuttavia anche questa più recente procedura non è esente da critiche: da un lato la stessa attenzione sulla concettualizzazione del costrutto produce un sensibile allontanamento dal sempre imprescindibile piano empirico e dall'altro lato non viene riconosciuto un netto carattere di generalità e quindi di estendibilità ad un ampio spettro di situazioni.

In questo contesto Finn e Kayande presentano quindi una ulteriore procedura, sintesi di quelle precedenti, che unisce l'approccio puramente empirico di Churchill a quello soprattutto razionale di Rossiter e che non trascurava l'importante aspetto di generalità che una scala deve possedere: nasce in questo modo la teoria di generalizzabilità multivariata (MGT).

2.5.2 L'approccio della teoria di generalizzabilità multivariata

La teoria di generalizzabilità multivariata si articola in due fasi: la prima riguarda l'aspetto della concettualizzazione e quello della conseguente confutazione con prove empiriche, la seconda invece riguarda un'applicazione pratica necessaria a produrre informazioni a supporto delle decisioni che un soggetto deve prendere.

La prima fase inizia con l'identificazione dell'oggetto che si va a misurare in tutte le sue componenti. Successivamente si esaminano tutte quelle caratteristiche che definiscono i requisiti

secondo i quali è possibile effettuare una generalizzazione: i soggetti valutatori, le condizioni di somministrazione della scala e la natura stessa della variabile che si sta studiando. A questo punto si crea il questionario attraverso degli opportuni item e lo si somministra ad un campione effettuando una selezione anche sugli aspetti di generalizzazione sopracitati. Raccogliendo i dati si ricercano eventuali riscontri empirici che possano validare quanto teorizzato all'inizio.

Dopo una serie di iterazioni della prima fase si passa alla seconda: è lo stadio in cui si prendono delle decisioni a partire dal supporto informativo proveniente da un insieme di misure selezionate che vengono impiegate.

Il vantaggio di questo approccio sta nel fatto che, effettuando un campionamento dell'universo delle condizioni di misura, si ottiene uno strumento che per sua natura è generale e generalizzabile a diversi ambiti posta una conferma empirica: la procedura C-OAR-SE in questo senso è decisamente all'opposto in quanto essa prevede che in ogni contesto venga creata un'apposita misura.

In questo modo la variabilità delle varie componenti delle condizioni di misura è acquisita "a monte" rendendo possibile la selezione degli aspetti più opportuni nella seconda parte dello studio.

2.5.3 Il modello matematico alla base della teoria di generalizzabilità multivariata

Il modello statistico alla base della teoria di generalizzabilità multivariata prevede che, prese due misure a e b di un costrutto, i livelli di punteggio ($X_{or,a}$ e $X_{or,b}$) ottenuti da un singolo valutatore siano la somma di una quantità corrispondente ad un punteggio medio universale (μ_a e μ_b) e ad un serie di quantità che tengono conto di quelle fonti che possono far variare il punteggio come la natura dell'oggetto (v_o e e_o), il soggetto rispondente (v_r e e_r), l'interazione tra oggetto e rispondente e altre fonti di errore ($v_{oe,r}$ e $e_{oe,r}$). Le seguenti equazioni sintetizzano il concetto:

$$X_{or,a} = \mu_a + v_o + v_r + v_{oe,r},$$

$$X_{or,b} = \mu_b + e_o + e_r + e_{oe,r}.$$

Il contributo di ogni fonte di errore in termini di varianza è espresso dalle seguenti matrici di varianza-covarianza:

$$\sum_o = \begin{bmatrix} \sigma_a^2(o) & \sigma_{ab}(o) \\ \sigma_{ab}(o) & \sigma_b^2(o) \end{bmatrix},$$

$$\sum_r = \begin{bmatrix} \sigma_a^2(r) & \sigma_{ab}(r) \\ \sigma_{ab}(r) & \sigma_b^2(r) \end{bmatrix},$$

$$\sum_{or,e} = \begin{bmatrix} \sigma_a^2(or, e) & \sigma_{ab}(or, e) \\ \sigma_{ab}(or, e) & \sigma_b^2(or, e) \end{bmatrix}.$$

La varianza in questo modello viene quindi scomposta nelle diverse cause che la determinano indipendentemente dal campionamento delle condizioni di misura.

Un'ultima quantità rilevante è la seguente correlazione che dà indicazioni sulla forza del legame tra le due diverse misure anche in questo caso in qualsiasi condizione di misura vengano campionate e fornisce anche un test per la validità:

$$\rho_{ab}(o) = \sigma_{ab}(o) / \sqrt{\sigma_a^2(o)\sigma_b^2(o)}.$$

2.5.4 Un esempio pratico

Il costrutto che viene qui preso ad esempio è quello della qualità del dettagliante che si articola nelle due dimensioni di qualità del servizio e qualità del prodotto. Per ciascuna di queste sono stati definiti tre item.

In questo contesto l'oggetto è rappresentato dai dettaglianti che vengono selezionati all'interno di un universo di venditori al dettaglio che trattano merci quali mattoni in una città di medie dimensioni. I soggetti valutatori invece vengono campionati tra quelle persone che hanno acquistato un bene negli stessi negozi della stessa città negli ultimi dodici mesi. Per verificare la validità convergente e divergente come pure per la validità del costrutto vengono utilizzate altre tre scale diverse che misurano la performance del negozio, l'attitudine verso il dettagliante e la fedeltà al punto vendita.

Per prima cosa i risultati indicano come le due dimensioni della variabile individuate siano corrette: la correlazione tra le due è infatti bassa.

Il calcolo delle correlazioni tra i singoli item mostra poi come, per quanto riguarda la dimensione della qualità del prodotto, le tre sue declinazioni abbiano un legame sensibile che esclude eventuali ridondanze; per quanto riguarda la dimensione della qualità del servizio, due item hanno una correlazione forte che può permettere in questo caso l'esclusione di uno dei due a patto poi che si esegua una ulteriore verifica empirica della nuova forma della scala.

È necessario sottolineare quest'ultimo aspetto, perché identifica in pieno l'approccio del metodo MGT. La concettualizzazione del costrutto infatti deve essere sempre supportata da prove empiriche che in un approccio C-OAR-SE non vengono cercate ma piuttosto vengono subordinate al giudizio di esperti: nella situazione specifica magari l'item non verrebbe eliminato ma riformulato se i giudici lo ritengono importante per cogliere al meglio il costrutto.

Da ultimo l'analisi del legame con altre scale di misura mostra convergenza con la performance del dettagliante, divergenza con l'attitudine e un legame con la fedeltà al punto di vendita.

La teoria di generalizzabilità multivariata qui esposta possiede dunque caratteristiche che la rendono migliore e preferibile ad altre procedure per lo sviluppo di scale di misura.

Essa infatti non è soggetta ai limiti di eccessiva concettualizzazione di Rossiter né di eccessivo empirismo di Churchill, ma è una perfetta sintesi dei loro punti di forza: precisa ricerca di item che possano cogliere al meglio il costrutto ma anche allo stesso tempo forte dipendenza dai riscontri pratici di quanto si teorizza.

L'approccio di generalizzazione che, tra le altre cose, è esso stesso una sorta di assicurazione sulla affidabilità di una misura e l'attenzione a tutti quegli aspetti che rientrano nel campo di quelle che vengono chiamate condizioni di misura (per cui non solo l'"OAR" di Rossiter) sono infine gli elementi che fundamentalmente caratterizzano questa teoria e la rendono uno strumento più raffinato e preciso di qualsiasi altro proposto fino a quel momento.

Capitolo 3: Studio di alcune variabili di marketing

3.1 Introduzione

I costrutti di interesse per il marketing sono, come è facile pensare, innumerevoli e dotati ciascuno di una cruciale importanza. È già stato sottolineato in precedenza come sia ad esempio utile e quasi indispensabile per un'azienda conoscere il livello di soddisfazione che un venditore nel campo industriale ha nei confronti del suo lavoro. La consapevolezza da parte delle imprese di dover porre attenzione a questi fenomeni è chiara oggi più che mai vista l'intensificazione della concorrenza. In questo capitolo vengono analizzate variabili ognuna delle quali di particolare interesse: il coinvolgimento del consumatore, l'orientamento al mercato di un'azienda e il taboo nella pubblicità.

3.2 Il coinvolgimento del consumatore

La misurazione del costrutto del coinvolgimento del consumatore è affrontata da J.L. Zaichkowsky in un articolo del 1985 apparso sul *Journal of Consumer Research*, dal titolo "*Measuring the Involment Construct*".

Il punto di partenza di questo studio è la constatazione della presenza di varie definizioni della variabile di marketing in questione e di conseguenti varie misure che a volte possono portare a conclusioni anche discordanti tra loro.

Con questo articolo Zaichkowsky cerca quindi di proporre una più corretta descrizione del concetto di coinvolgimento e una migliore scala di misura che goda delle solite proprietà di validità ed affidabilità, anche prendendo spunto dai precedenti studi che hanno trattato la materia in questione.

3.2.1 Il concetto del coinvolgimento

La natura del coinvolgimento è riconducibile alle tre seguenti aree: area personale, che ha a che fare con gusti, interessi e valori del consumatore; area fisica, legata alle caratteristiche proprie dell'oggetto; area circostanziale, che riguarda aspetti situazionali che possono essere temporanei.

Gli oggetti che possono creare un coinvolgimento più o meno intenso del consumatore sono i seguenti: la pubblicità, il prodotto e il processo di acquisto.

Alcune teorie precedenti suppongono che il consumatore (sostanzialmente razionale) affronti sempre una fase preliminare di accurata raccolta di informazioni prima dell'acquisto di un prodotto. Tuttavia altri studi dimostrano come il coinvolgimento del consumatore possa invece variare ed appartenere fondamentalmente a due categorie: quella di alto coinvolgimento e quella di basso coinvolgimento.

Nell'articolo di Zaichkowsky è in base a quest'ultima visione che viene sviluppata la procedura di costruzione della scala di misura.

3.2.2 Il processo di sviluppo della misura

Il tipo di scala che in questo contesto è considerato come il migliore al fine di cogliere il concetto di coinvolgimento è quello semantico bipolare: esso consiste in una serie di item bipolari, ognuno dei quali misurato con una scala a sette punti.

I vantaggi nell'usare questo metodo stanno nella sua versatilità per quanto riguarda gli oggetti a cui esso è applicabile e nella facilità di somministrazione e calcolo dei punteggi.

Il processo di sviluppo della misura si articola attraverso i seguenti passaggi: definizione del concetto, generazione degli item e verifica della validità dei contenuti, studio dell'affidabilità degli item generati e ulteriore selezione di questi, nuovo accertamento della validità dei contenuti degli item appena selezionati, verifica della validità attraverso il legame con un altro criterio, accertamento della validità del costrutto.

Risultano evidenti le somiglianze con il lavoro di Churchill e Gilbert (1979) nella realizzazione di questa scala.

- *Definizione del costrutto*

In questo studio il concetto di rilevanza personale è alla base della definizione di coinvolgimento: pubblicità rilevante in quanto il consumatore è stimolato da essa; prodotto rilevante in merito ai bisogni che esso soddisfa; decisione d'acquisto rilevante e in quanto tale degna di attenzione.

In generale, al di là di quale sia l'oggetto in questione (pubblicità, prodotto, decisione d'acquisto), si può dire che un alto coinvolgimento è segno di rilevanza personale.

La definizione di coinvolgimento che viene qui adottata (che si rifà anche a precedenti lavori) è dunque la seguente: "La rilevanza percepita di un oggetto da parte di una persona, basata sui suoi bisogni, valori ed interessi".

- *Generazione degli item e validità dei contenuti*

Per rappresentare il concetto di coinvolgimento sono state selezionate 168 coppie di termini bipolari (importante-non importante, interessante-non interessante, ecc.), ognuna delle quali rappresentante un item; è quindi stata valutata la loro capacità di cogliere la variabile del coinvolgimento, scartando per prima cosa gli item meno rappresentativi ed eseguendo poi un'ulteriore accurata valutazione delle coppie più appropriate.

Il compito di selezione è toccato a dei giudici esperti, che hanno classificato i 168 item nelle categorie "chiaramente rappresentativo", "in qualche modo rappresentativo" e "non rappresentativo".

La valutazione è stata effettuata per i tre tipi di oggetti sopra descritti, ovvero pubblicità, prodotto e decisione d'acquisto. Da questa fase di scrematura sono usciti 43 item, ulteriormente ridotti a 23 da altri cinque nuovi giudici. Infine ne sono stati aggiunti 7 per raggiungere un totale di 30, ritenuto non sufficiente il numero di item ottenuti dalla seconda selezione.

- *Affidabilità interna della scala*

Il successivo passaggio è dedicato all'accertamento della coerenza interna, ovvero all'analisi della correlazione tra item.

Vengono scelti due tipi di prodotti diversi (orologi e scarpe da ginnastica) e, ad un campione di 150 studenti, viene somministrata la scala (a metà un prodotto, all'altra metà l'altro).

Per entrambe le categorie i risultati mostrano una correlazione interna maggiore o uguale a 0,5 per 26 item, e un alfa allo 0,95.

Sono stati scartati poi sei item che presentavano correlazione con il totale relativamente bassa: un'analisi fattoriale ha escluso che gli item scartati potessero essere ricondotti ad un'unica dimensione soggiacente (e quindi nessun singolo item ad essa associato può essere aggiunto).

- *Test-retest per l'affidabilità*

A questo punto i 24 item sono stati testati su un nuovo campione (ancora studenti) e su quattro nuove categorie di prodotti (calcolatrici, colluttori, vino rosso e cereali per la colazione) per due volte a distanza di tre settimane.

I risultati che si sono ottenuti sono i seguenti: un coefficiente di correlazione di Pearson tra i due risultati pari a 0,90 e una correlazione tra identici item compresa tra 0,31 e 0,93.

Dopo l'accertamento dell'affidabilità interna e di quella test-retest si è pervenuti quindi ad una scala definitiva di 20 item (sono stati eliminati i 4 che avevano una correlazione tra il punteggio della prima somministrazione e quello della seconda inferiore a 0,60), che viene denominata "Personal Involvement Inventory" (a dieci item presi a caso è stato ribaltato il punteggio per una questione di bilanciamento).

- *Validità dei contenuti (2)*

Il Personal Involvement Inventory è stato poi sottoposto ad una ulteriore verifica della validità dei contenuti.

Ad un nuovo campione di studenti è stata somministrata la scala (per tre differenti prodotti anche in questo caso) e successivamente è stato chiesto loro attraverso una domanda a risposta aperta di motivare la loro valutazioni sui prodotti.

In base ai risultati ottenuti, i rispondenti sono stati classificati prima in base al punteggio ottenuto (basso, medio, alto), poi in base alle risposte alla domanda aperta, divisi da due giudici esperti (ignari dei punteggi ottenuti) nelle categorie di "basso coinvolgimento", "medio coinvolgimento", "alto coinvolgimento".

Dai dati della divisione in categorie ci si è accorti di una significativa somiglianza tra la categorizzazione dei punteggi e quella delle domande a risposta aperta; ciò non è altro che una prova a supporto della validità del contenuto cercata.

- *Verifica della validità attraverso il legame con un altro criterio*

Verificando la validità legata ad un altro criterio si vanno a confrontare i punteggi della scala sviluppata con quelli di un altro metodo di misurazione della stessa caratteristica; in questo caso il criterio scelto è quello della classificazione dei prodotti in basso o alto coinvolgimento (21 prodotti

oggetto di studi precedenti, una scala a otto punti che va da “estremamente non importante nella mia vita” a “estremamente importante nella mia vita”).

Somministrato il PII ad un campione di 47 studenti che dovevano esprimersi su 4 dei 21 prodotti, si sono confrontate le due valutazioni e si è ottenuta una significativa convergenza tra i due differenti risultati.

- *Validità del costrutto*

Studiando la validità del costrutto si va ad indagare sulla natura di ciò che si vuole catturare attraverso la scala di misura; ci si chiede quale sia la sua vera coerenza, si cerca di capire se si stia analizzando un concetto ben chiaro e definito e non semplicemente qualcosa di artificialmente creato dallo studioso.

Il passo iniziale è quindi quello di scoprire cosa ci sia alla base della differenza tra alto e basso coinvolgimento.

Facendo riferimento a vari studi teorici presenti nella letteratura, Zaichkowsky stila il seguente elenco di possibili motivazioni (qui da una prospettiva di basso coinvolgimento):

- Mancanza di una ricerca attiva di informazioni sulle varie marche;
- Scarso confronto tra gli attributi dei prodotti;
- Percezione di somiglianze tra marche diverse;
- Nessuna preferenza per una marca in particolare.

Per capire come e quanto siano presenti questi fattori viene costruita una misura formata dalle seguenti affermazioni, ciascuna valutabile secondo una scala a 7 punti i cui estremi sono rappresentati da “fortemente in disaccordo” e “fortemente d’accordo”:

- Sarei interessato a leggere informazioni riguardanti il modo in cui il prodotto è stato fatto;
- Sarei interessato a leggere articoli in cui altri consumatori valutano il prodotto;
- Ho confrontato le caratteristiche del prodotto tra le differenti marche;
- Penso che vi siano differenze significative tra prodotti di diverse marche;
- Preferisco una marca in particolare.

Questa scala e il PII sono state somministrate ad un campione composto da 57 persone (non tutte hanno accettato, alcune non hanno completato il questionario) e sono state scelte tre categorie di prodotti (caffè istantaneo, televisione a colori, detersivo per lavatrice).

I rispondenti sono stati quindi classificati in base ai punteggi del PII in tre categorie (basso, medio, alto coinvolgimento) come è avvenuto nella fase in cui si cercava di accertare per la seconda volta la validità dei contenuti.

Lo scopo del successivo studio dei dati raccolti è sostanzialmente quello di vedere se i soggetti che hanno un basso punteggio nel PII hanno fornito valutazioni diverse rispetto a quelli che hanno invece un punteggio alto. A tal fine sono stati calcolati coefficienti di correlazione di Pearson tra i risultati delle due misure, coefficienti alfa, t-test.

Da questa analisi è risultata convergenza tra ciò che già si sapeva da precedenti studi sul concetto di coinvolgimento e tra i vari indici calcolati. Per citare un esempio, prendendo come riferimento le prime due affermazioni della seconda scala, a livello teorico un relativo alto punteggio dovrebbe corrispondere ad un alto coinvolgimento: i dati empirici trovati lo confermano e ciò vale se si vanno a considerare tutti i prodotti e per tutte le altre affermazioni.

L'ultimo passo da compiere è quello di verificare l'unidimensionalità della variabile attraverso l'analisi fattoriale.

In definitiva gli esiti positivi dell'indagine compiuta in questa fase non sono altro che la prova cercata della validità del costrutto: ciò che si va a misurare, ovvero il coinvolgimento, non è dunque una mera "invenzione" dello studioso quanto piuttosto un vero costrutto esistente e misurabile.

Scorrendo attraverso le diverse fasi del processo di sviluppo del PII con il fine misurare il coinvolgimento del consumatore presentato da Zaichkowsky è facile quanto importante notare come esso sia costituito da una serie di rigorosi e precisi accertamenti delle varie proprietà (validità dei contenuti, affidabilità, coerenza interna, validità del costrutto); si è cercato di unire le conoscenze teoriche elaborate a partire da quanto era presente in letteratura con quanto risultava dalle valutazioni dei soggetti che erano chiamati a rispondere al questionario (o anche degli esperti che in uno stadio iniziale dovevano dare dei giudizi sulla validità dei contenuti degli item) e dai conseguenti indici calcolati.

Questo metodo riesce in sostanza a raggiungere lo scopo che in partenza è stato prefissato, ovvero quello di cogliere la variabile del coinvolgimento del consumatore in maniera efficace, carpando a pieno le sue sfumature.

3.3 I taboo nella pubblicità

Nell'ambito delle teorie dello sviluppo di scale di misura applicate ad un particolare costrutto rientra *"Taboo in advertising: conceptualization, scale development and analysis of its communication effects"*, un articolo scritto da Ouidade Sabri-Zaaraoui e presentato durante la 34-esima conferenza internazionale di Marketing "La Londe" nel 2007.

Questo articolo va a riempire un vuoto presente nella letteratura di marketing: mai infatti era stato condotto uno studio approfondito e curato sull'argomento dei taboo nella pubblicità, sia per quanto riguarda la sua concettualizzazione sia per quanto riguarda una sua misurazione.

Uno studio questo che si rileva molto utile e fondamentale, visto il fatto che da parte di molte aziende l'uso dei taboo nella pubblicità è una pratica ormai assodata: la concorrenza sempre più agguerrita ha messo infatti le industrie nella condizione di sfruttare al massimo i vari argomenti taboo per cercare di attirare più consumatori possibili.

L'attenzione è quindi posta da Sabri-Zaaraoui sui seguenti obiettivi: raggiungere una buona concettualizzazione del taboo, sviluppare una scala di misura del concetto che goda delle proprietà di affidabilità e validità e infine valutare gli effetti che si ottengono attraverso l'utilizzo dei taboo, in particolare descrivere la dinamica che può trasferire il giudizio positivo o negativo della pubblicità sul prodotto, sulla marca e sull'azienda.

3.3.1 Il concetto di taboo

La materia dei taboo è stata tradizionalmente affrontata sin dai primi anni del ventesimo secolo da sociologi, antropologi e psicoanalisti; gli studiosi di marketing, come è facile pensare, si sono avvicinati ben poco a questo argomento.

Tutti gli studiosi sono concordi nel definire il taboo prima di tutto come un qualcosa di proibito, la cui realizzazione è da evitarsi per il bene della società; in secondo luogo gli si dà una natura non razionale quanto piuttosto radicata in norme stabilite dalla comunità che si sono rafforzate nel corso della storia di una civiltà; da ultimo, vengono identificati due tipi di taboo, uno relativo a comportamenti che non possono essere tenuti, l'altro relativo invece a cose che non possono essere dette.

Un esempio di taboo può essere quello della morte: esso ha a che fare contemporaneamente sia con la sfera comportamentale (la proibizione dell'omicidio) che con quella del linguaggio (evitare di parlare in pubblico della morte).

3.3.2 Il taboo e la pubblicità

Nella pubblicità di un prodotto il taboo può essere presente a due livelli diversi: può essere il prodotto stesso ad essere in qualche modo un taboo, o può essere la pubblicità ad usare temi ed immagini taboo che non hanno comunque a che fare con ciò che viene pubblicizzato. Nel caso di questo studio ci si concentra su quest'ultimo ruolo, ovvero ci si focalizza sull'uso di taboo che sono incongruenti con ciò che si va a promuovere, per capire quali tipi di norme vengono violate.

Il taboo fa parte di una esplicita strategia di provocazione adottata per accattivarsi il consumatore; lo scopo è creare una forte reazione nei potenziali acquirenti (ma non solo) utilizzando i più svariati temi quali il suicidio, l'omicidio, il sesso nelle sue devianze, e tanti altri. I prodotti legati a questi tipi di promozioni sono quelli tra i più ordinari: vestiti, scarpe, bibite ne sono alcuni esempi.

Bisogna precisare poi che provocazione e taboo sono concettualmente distinti: prima di tutto quest'ultimo è un mezzo per ottenere la prima ed, in secondo luogo la provocazione può anche trasgredire cose che non rappresentano taboo come, ad esempio, valori individuali o collettivi. Per questi motivi i taboo hanno una loro precisa area di ricerca.

Come già detto, scarsi sono i lavori sulla concettualizzazione e la misura dei taboo nella pubblicità, per questo è necessario sviluppare una misura che ben riesca a cogliere le due dimensioni del costrutto, ovvero quella relativa a violazioni di norme comportamentali e quella relativa a violazioni di norme conversazionali.

3.3.3 Lo sviluppo della scala

La scala che viene sviluppata in questo contesto si basa sulle teorie di Churchill (1979) e Rossiter (2002) e si propone di misurare il grado di taboo percepito in una pubblicità da parte di un ricevente.

Essa si articola in due fasi: la prima di costituzione vera e propria della scala, che comprende la selezione iniziale di item e delle pubblicità, una purificazione della scala attraverso un'analisi fattoriale esplorativa ed un'analisi fattoriale confermativa; la seconda invece di verifica della varie proprietà di cui deve godere una misura, ovvero si cerca di accertare l'affidabilità dello strumento, la validità dei contenuti, le validità convergente e divergente.

- *Item iniziali e selezione delle pubblicità*

Il punto di partenza è quello della selezione degli item e della scelta delle pubblicità che verranno sottoposti ai soggetti valutatori.

Per quanto riguarda il primo compito è stato creato inizialmente un set di 20 item attraverso lo studio della letteratura disponibile e attraverso 22 interviste in profondità; successivamente 8 esperti si sono espressi sulla capacità degli item selezionati di cogliere il costrutto nelle sue due dimensioni (una sorta di preliminare verifica della validità dei contenuti), scremando il set iniziale a 14 item (alcuni dei quali riveduti rispetto alla formulazione iniziale).

Per quanto riguarda invece il secondo compito, si è partiti mostrando 16 pubblicità che rappresentano i due tipi di taboo (quello comportamentale e quello conversazionale) ad un campione di 30 persone (età media 38) e ponendo loro le seguenti domande misurate da una scala di Likert a 5 punti: “Questa pubblicità rappresenta un taboo per lei?” e “Secondo lei, cosa suggerisce questa pubblicità?”. Dopo la somministrazione sono state selezionate quattro pubblicità, in due delle quali è stato percepito un leggero taboo mentre nelle altre due un forte taboo; per ognuna di queste è stato creato infine un questionario, rappresentante dunque un singolo taboo.

- *Analisi fattoriale esplorativa per la purificazione della scala*

L’analisi fattoriale esplorativa, suggerita ad esempio da Gerbing ed Anderson (1988), è un ottimo strumento per identificare le dimensioni di un costrutto e al tempo stesso per scremare ulteriormente il set di item conseguito fino a questo punto.

Quest’analisi è stata compiuta a partire dalle risposte da parte di una campione di convenienza composto da 180 persone di età media pari a 31, maschi poco più della metà.

Si è proceduto attraverso l’analisi delle componenti principali, la quale ha selezionato 6 item; da qui si è proceduto con lo stimare il contributo alla varianza totale di ciascuna delle due dimensioni, ed è risultato essere complessivamente pari al 68%.

L’analisi effettuata insomma non ha fatto che dimostrare la validità di ciò che si è precedentemente teorizzato: il costrutto del taboo nella pubblicità è costituito da due componenti, una riguardante taboo comportamentali e l’altra riguardante taboo conversazionali.

- *Analisi fattoriale confermativa*

A questo punto, tramite la raccolta di dati ed ad un’altra analisi fattoriale, si cerca di trovare una conferma di quanto emerso dalla precedente analisi fattoriale esplorativa.

È stato selezionato un campione di 213 persone di età compresa tra 30 e 59, in maggioranza donne (56%), e sono state utilizzate le stesse 4 pubblicità adottate all'inizio.

Una serie di indici, tra cui il chi-quadro che verifica l'ipotesi nulla di correttezza del modello, danno conferma di quanto si voleva accertare: ciò che è risultato è in pieno accordo con l'output dell'analisi esplorativa.

- *Verifica delle proprietà*

Per quanto riguarda l'affidabilità, i due coefficienti alfa di Cronbach calcolati per ogni dimensione del costrutto indicano una buona coerenza interna.

La validità dei contenuti invece è stata accertata nella fase iniziale di selezione degli item attraverso il giudizio di otto soggetti esperti: tre professori di marketing, due pubblicitari che lavorano in note agenzie e tre consumatori.

Infine attraverso il confronto della varianza media estratta delle due dimensioni con la radice quadrata della correlazione delle stesse si dimostra la validità divergente.

3.3.4 Effetti dell'uso di taboo nella pubblicità

La costruzione della scala è stata ultimata, così come le diverse proprietà che essa deve avere sono state accertate: resta da analizzare l'effetto che può derivare dall'uso del taboo nella pubblicità.

L'origine del ragionamento che soggiace a questa fase si fonda su due ipotesi che sono state confutate da studi precedenti: la prima è che le donne percepiscono di più degli uomini la presenza di un taboo in una pubblicità, la seconda invece che le persone anziane hanno più sensibilità di quelle giovani nel riconoscere un taboo nella pubblicità. Sesso ed età insomma sono due variabili discriminanti nella valutazione del grado di presenza di un taboo.

I dati raccolti nelle due somministrazioni danno un'ulteriore prova della veridicità delle due ipotesi: effettuando t-test si scopre una significativa differenza sia tra i punteggi medi femminili e maschili sia tra i punteggi delle persone vecchie e giovani.

Appurato ciò, si passa a considerare i veri e propri effetti dell'uso del taboo, ipotizzando questa serie consequenziale di affermazioni: il grado percepito di taboo ha un effetto negativo sull'attitudine verso la pubblicità; il grado percepito di taboo è positivamente e fortemente correlato all'attitudine verso la marca; il grado percepito di taboo è positivamente e fortemente correlato all'intenzione di acquisto. In altre parole, il grado della percezione di un taboo in un a pubblicità può avere un'influenza finale sulle vendite del prodotto pubblicizzato.

Anche in questo caso le informazioni provenienti dalla somministrazione dei questionari provano la autenticità delle ipotesi appena descritte.

La concettualizzazione, la misurazione e la valutazione degli effetti del taboo nella pubblicità è infine giunta al termine; si è provata la bidimensionalità del fenomeno, si sono verificate le necessarie condizioni di affidabilità e validità della scala, sono stati messi in risalto alcune considerazioni, quali il diverso effetto di una pubblicità legato all'età e al sesso del ricevente del messaggio pubblicitario.

Questo studio deve anche però scontrarsi con dei limiti: la ricerca è stata condotta in un contesto relativamente ristretto (quello francese), e la validità esterna non è molto solida.

Considerato tutto ciò, la scala qui presentata può essere ritenuta un strumento utile per quei pubblicitari che decidono di servirsi dei taboo; lungi dall'essere la miglior misura ottenibile, essa resta comunque un ottimo punto di partenza per ulteriori studi che possano renderla ancora migliore.

3.4 L'orientamento al mercato di un'impresa

Tra gli studi applicativi di sviluppo di scale di misura si annovera anche "*MARKOR: A Measure Of Market Orientation*" pubblicato nel 1993 sul *Journal of Marketing Research* realizzato da A.K. Kohli, B.J. Jaworski e A. Kumar.

Il costrutto di marketing analizzato è quello dell'orientamento al mercato di un'impresa, una variabile sulla quale non era mai stata creata prima di questa opera una scala di misura generale ma erano invece presenti una serie di studi costruiti sui singoli casi specifici che non potevano quindi essere estesi oltre quell'ambito ristretto.

Come ben sottolineato dagli autori questo concetto è di fondamentale importanza per le attività di marketing che un'impresa svolge. Ecco quindi che questo studio quanto mai necessario si propone di trattare in maniera precisa e sistematica il problema della misura del costrutto dell'orientamento al mercato creando uno strumento di misurazione universale dotato delle proprietà psicometriche opportune prendendo come punto di riferimento gli strumenti teorici basilari forniti dalle acquisizioni di Churchill e Peter (1979) e di Anderson (1988).

3.4.1 La realizzazione della scala

La costruzione della scala si articola secondo una serie di passaggi che seguono sostanzialmente le indicazioni fornite da Churchill (1979): specificazione del dominio, generazione degli item, purificazione della misura, raccolta dei dati e analisi.

- *Specificazione del dominio*

Nella specificazione del dominio il primo passo da compiere è quello di eseguire una ricognizione di quanto è già presente nella letteratura.

Nel caso dell'orientamento al mercato esistono pubblicazioni che concettualizzano il costrutto con sensibili differenze, ma in generale si possono riassumere i tratti comuni ricorrenti: il punto di vista del consumatore, il marketing coordinato e la redditività.

Per ottenere una specificazione ancora più precisa e completa sono stati interpellati dei manager e si è arrivati all'identificazione di tre componenti che formano l'orientamento al mercato: la generazione di informazioni tesa a capire le dinamiche che stanno alla base dei bisogni e delle preferenze dei consumatori, la diffusione delle informazioni all'interno di un'organizzazione e la risposta dell'azienda in relazione a quanto ottenuto dalla raccolta e dalla diffusione delle informazioni.

- *Generazione degli item*

La generazione degli item è svolta a partire dalle interviste realizzate ai manager creando 25 item in grado di cogliere tutte e tre le componenti della variabile di marketing in analisi.

- *Purificazione della scala*

La purificazione della scala si svolge attraverso la somministrazione di tre test preliminari. Il primo, che comporta l'esclusione di 4 item, viene affidato a 27 manager di cui solo una parte appartenente alla sfera del marketing. L'eliminazione degli item è stata decisa in base al loro grado di ambiguità e di non chiarezza, tenendo anche conto della correlazione del punteggio di un ciascun item con il punteggio totale e della dipendenza di un item a più di una componente. Il secondo pretest, che ha coinvolto sette esperti accademici chiamati ad esprimersi sulla capacità di cogliere il dominio, ha portato alla modifica di alcuni item, all'eliminazione e all'aggiunta di altri, per arrivare ad una scala di 32 item. Il terzo ed ultimo test, che ha coinvolto sette manager ai quali è stato chiesto di compilare e commentare il questionario, non ha portato modifiche sostanziali.

- *Raccolta dei dati*

Per la raccolta dei dati è stato scelto un campione di 500 manager in ambito di marketing selezionati da una lista fornita dalla American Marketing Association ai quali è stato inviato per posta il questionario ottenuto nella fase precedente; il tasso di risposta è stato circa del 47% per un totale di 230 rispondenti. Una seconda ulteriore raccolta dati ha individuato un campione di manager appartenenti all'area marketing e non scelti tra 13 imprese appartenenti al Marketing Science Institute e un campione di amministratori delegati presi dalla "Dun and Bradstreet top 1000 U.S. firms".

- *Analisi dei dati raccolti*

L'analisi dei dati raccolti si sviluppa in tre fasi: nella prima si usano le informazioni ottenute attraverso il primo campione per individuare e scartare quegli item inadatti a descrivere la variabile, nella seconda si usano i dati raccolti successivamente per verificare alternative strutture del costrutto ed anche per identificare quelle componenti a cui si riferiscono gli item appena scartati, nella terza si verifica infine la validità predittiva della scala calcolando la correlazione delle componenti del costrutto con altre variabili significative.

L'analisi della prima fase ha portato a preferire un modello in cui è presente oltre ai tre fattori già individuati e correlati un fattore generale che include i tre aspetti e a scartare 12 item in base a criteri quali il loro maggiore contributo alla varianza d'errore stimata.

Nella seconda fase si cercano ancora dei possibili fattori soggiacenti il costrutto stimando una serie di modelli che sono la combinazione di un certo sottoinsieme dei seguenti tre tipi di fattori: uno generale di orientamento al mercato comune a tutte le variabili aggiuntive che si prendono in considerazione (40 in tutto), un altro per ciascuno delle tre componenti individuate in precedenza ed uno per tipo di rispondente (manager di area marketing o meno). Questa analisi ha come principale risultato la riscrittura di alcuni item che sono risultati essere formulati in maniera non adeguata.

Per la terza fase i costrutti scelti per il confronto sono stati i seguenti: l'enfasi del top management sull'orientamento al mercato, i conflitti tra diverse funzioni, l'impegno organizzativo degli impiegati, l'accertamento delle performance dell'azienda, tutti misurati attraverso più di un item, e una sesta variabile globale di orientamento al mercato associata ad un singolo item. Calcolando la correlazione dei costrutti con le singole componenti dell'orientamento al mercato sono infine state trovate prove della validità che si andava cercando.

La realizzazione della scala MARKOR per la misura dell'orientamento al mercato di un'impresa è quindi conclusa. Partendo dall'individuazione delle componenti della variabile e passando attraverso la costruzione della scala e la verifica delle proprietà si è ottenuto un buon strumento in grado di misurare il costrutto oggetto di studio.

Infine non mancano tuttavia spunti di particolare interesse che riguardano possibili sviluppi e miglioramenti della scala realizzata: una revisione attraverso ad esempio l'inserimento di nuovi item o la riverifica delle proprietà di validità, un confronto con altri strumenti già esistenti che trattano con sfumature diverse lo stesso costrutto, o ancora la semplice applicazione della stessa per farne risaltare limite e potenzialità inesprese.

BIBLIOGRAFIA

Gilbert A. Churchill, Jr., Neil M. Ford e Orville C. Walker, Jr., “*Measuring the Job Satisfaction of Industrial Salesmen*”, *Journal of Marketing Research* (1974).

Peter, J. Paul, “*Reliability: A Review of Psychometric Basics and Recent Marketing Practices*”, *Journal of Marketing Research* (1979).

Gilbert A. Churchill, Jr. , “*A Paradigm for Developing Better Measures of Marketing Constructs*”, *Journal of Marketing Research* (1979).

Judith L. Zaichkowsky, “*Measuring the Involvement Construct*”, *the Journal of Consumer Research* (1985).

David W. Gerbing e James C. Anderson, “*An Updated Paradigm for Scale Development Incorporating Unidimensionality and Its Assessment*”, *Journal of Marketing Research* (1988).

Ajay K. Kohli, Bernard J. Jaworski e Ajith Kumar, “*MARKOR: A Measure of Market Orientation*”, *Journal of Marketing Research* (1993).

Adamantios Diamantopoulos e Hedi M. Winklhofer, “*Index Construction with Formative Indicators: An Alternative to Scale Development*”, *Journal of Marketing Research* (2001).

John R. Rossiter, “*The C-OAR-SE procedure for scale development in marketing*”, *International Journal of Research in Marketing* (2002).

Adam Finn, Ujwal Kayande, “*How fine is C-OAR-SE? A generalizability theory perspective on Rossiter’s procedure*”, *International Journal of Research in Marketing* (2005).

Ouidade Sabri-Zaaraoui , “*Taboo in advertising: conceptualization, scale development and analysis of its communication effects*”, 34-esima conferenza internazionale di Marketing “La Londe” (2007).