



UNIVERSITÀ  
DEGLI STUDI  
FIRENZE

## FLORE

# Repository istituzionale dell'Università degli Studi di Firenze

### **La misurazione nella ricerca sociale. Teorie, strategie, modelli**

Questa è la Versione finale referata (Post print/Accepted manuscript) della seguente pubblicazione:

*Original Citation:*

La misurazione nella ricerca sociale. Teorie, strategie, modelli / F. MAGGINO. - ELETTRONICO. - (2004), pp. 1-99.

*Availability:*

The webpage <https://hdl.handle.net/2158/306284> of the repository was last updated on

*Publisher:*

FIRENZE UNIVERSITY PRESS, ARCHIVIO E-PRINTS

*Terms of use:*

Open Access

La pubblicazione è resa disponibile sotto le norme e i termini della licenza di deposito, secondo quanto stabilito dalla Policy per l'accesso aperto dell'Università degli Studi di Firenze (<https://www.sba.unifi.it/upload/policy-oa-2016-1.pdf>)

*Publisher copyright claim:*

La data sopra indicata si riferisce all'ultimo aggiornamento della scheda del Repository FloRe - The above-mentioned date refers to the last update of the record in the Institutional Repository FloRe

(Article begins on next page)

## 4. AFFIDABILITA' E VALIDITA'

### 4.1 LA VALIDITÀ

Come abbiamo visto la validità rappresenta la capacità di una procedura di misurazione di misurare ciò che si intende misurare. E' possibile distinguere diversi tipi di validità cui corrispondono metodi diversi di verifica; vediamoli schematicamente:

VALIDITÀ	Soft (translation validity)	L'operationalizzazione rappresenta un buon riflesso del costrutto	a. <i>face validity</i> : apparente, logica, concettuale
			b. <i>content validity</i> : di contenuto, intrinseca, circolare, rilevante, rappresentativa
			c. <i>construct validity</i> : della struttura, di costrutto, teorica, caratteristica, fattoriale
	Hard (criterion related validity)	L'operationalizzazione funziona in modo predittivo rispetto ad altre operationalizzazioni	d. <i>criterion validity</i> : predittiva, empirica, statistica, correlata al criterio
			e. <i>concurrent validity</i> : concorrente
			f. <i>convergent validity</i> : convergente
			g. <i>discriminant validity</i> : discriminante

- Face Validity**: è determinata dalla significatività apparente ed esteriore che una misura presenta, da cui il nome; per valutarla si richiedono giudizi di esperti relativamente alla validità che sembrano avere le misure.
- Content Validity**: una misura ha validità di contenuto quando i suoi indicatori rappresentano in modo accurato l'universo di contenuto misurato; anche questo tipo di validità richiede il giudizio di esperti. Se, per esempio, una prova di verifica della preparazione di storia di un gruppo di studenti è composta solo da domande riguardanti le Guerre di Indipendenza difficilmente potrà essere riconosciuta valida nel suo contenuto.
- Construct Validity**: è data dal livello in cui gli indicatori misurano accuratamente i costrutti teorici che interessa misurare; essa è verificata praticamente attraverso la misura della correlazione tra un indicatore ed altri indicatori secondo particolari modelli teorici. Poniamo, per esempio, di voler costruire uno strumento che misuri le tendenze razziste di un particolare gruppo di individui; secondo l'ipotesi, l'indice sarà una buona misura di razzismo, ovvero avrà validità di costrutto, se risulterà essere inversamente correlato con l'autostima; per poter procedere in questo modo è necessario assumere che la teoria sia corretta, ovvero che esista una relazione tra autostima e razzismo.
- Criterion Validity / Predictive Validity**: rappresenta la capacità di uno strumento di fare previsioni accurate; la verifica di tale validità è fatta a partire dalla sua adeguatezza nel predire un criterio esterno, da cui il nome.
- Concurrent Validity**: è determinata osservando quanto lo strumento correla con altri strumenti che il ricercatore ritiene validi nel misurare la stessa caratteristica; l'osservazione di una relazione statisticamente significativa è considerata verifica della validità. Se, per esempio, si vuole verificare se una particolare misura rappresenti uno strumento valido per misurare l'autostima, è possibile correlare i punteggi con essa ottenuti dai soggetti, con i punteggi ottenuti dagli stessi soggetti con una diversa scala ritenuta valida nel misurare l'autostima; un'alta correlazione evidenzierà la validità del nuovo test o, almeno, che i due test stanno misurando la stessa cosa.
- Convergent Validity**: è determinata confrontando e correlando i punteggi ottenuti con la misura da validare con quelli ottenuti con la misura di un altro costrutto, teoricamente legato al primo. La possibilità di verificare la validità convergente dipende quindi dall'esistenza di costrutti, e relative misure, legati con quello misurato.

- g. *Discriminant Validity*: è speculare alla validità convergente; essa è alta quando la misura da validare non correla con le misure di altri costrutti, teoricamente distinti dal primo. Supponiamo, per esempio, di voler predisporre un questionario strutturato per misurare il costrutto *patriottismo* in relazione al quale si può ipotizzare che negli individui esista un atteggiamento "reale" e un atteggiamento che risponda a una certa *desiderabilità sociale*; in questo caso è possibile predisporre un gruppo di indicatori che misurano il costrutto *risposte socialmente desiderabili*; se le due misure (*patriottismo* e *desiderabilità sociale*) non risultassero correlate, la misura di patriottismo avrebbe più validità discriminante ovvero sarebbe non correlata ad una misura di qualcosa con la quale non dovrebbe essere correlata se fosse valida.

### 4.1.1 Modelli di verifica

Uno dei momenti più difficili nel processo di sviluppo di una misura è quello dell'interpretazione dei punteggi ottenuti dalla misurazione. La registrazione di una significativa affidabilità rivela solo che lo strumento misura veramente qualcosa ma non dà alcuna informazione sulla natura di ciò che si sta misurando. La difficoltà deriva dal fatto che la validazione può essere verificata all'interno di un sistema di relazioni ipotizzate tra il costrutto di interesse e altri costrutti; tali relazioni possono essere di causa, effetto o di correlazione. Le misure empiriche sono utilizzate per verificare tali ipotesi; il supporto empirico per le ipotesi comporta la validità della misura. Le prove di validità implicano la verifica simultanea dell'ipotesi riguardante i costrutti e lo strumento.

Come si è già detto, molti costrutti nelle scienze sociali rappresentano astrazioni teoriche che non esistono nell'esperienza reale; per tale motivo gli indicatori che li rappresentano devono essere sottoposti a verifica di validità. La validazione di una misura può quindi essere vista come la verifica di una teoria. Il processo di validazione non può essere affrontato prima che non sia stato portato a termine quello di verifica dell'affidabilità.

#### 4.1.1.1 Le strategie

Approfondiremo a questo punto le diverse strategie di verifica dei tre principali tipi di validità che più interessano la ricerca sociale e che corrispondono alle tre principali funzioni rispetto all'utilizzo di uno strumento di misurazione:

TIPO DI VALIDITA'	FUNZIONE
1. PREDITTIVA	stabilire una relazione statistica con una particolare variabile
2. DI CONTENUTO	rappresentare/rilevare un universo di contenuto specifico
3. DELLA STRUTTURA o TEORICA	misurare particolari tratti psicologici, rapporti sociali, ecc.

Anche se i tre tipi di validazione sono presentati separatamente nella pratica essi tendono ad essere complementari in quanto ciascuno di essi è sostenuto o riceve informazioni dall'altro.

#### *La validità predittiva*

Il modello statistico e la logica utilizzati per studiare e verificare la validità predittiva sono relativamente semplici, anche se dal punto di vista applicativo il procedimento può risultare piuttosto complesso.

Alla base del modello di verifica vi è il confronto tra la misura da validare ed altre misure dette *criteri*<sup>1</sup>. Tale approccio prevede e richiede la definizione e la verifica di una ipotesi riguardante le

<sup>1</sup> Tra i diversi tipi di criterio esistenti, secondo alcuni autori, è possibile distinguere le seguenti tipologie:

1. *product criteria*, rappresentati da variabili legate a particolari prestazioni e risultati di alcune specifiche attività (per esempio numero di lettere battute a macchina, numero di scatole ispezionate, ecc.);
2. *action criteria*, che riguardano l'attività stessa (per esempio misurazioni di particolari prestazioni: velocità o numero di errori nell'eseguire qualcosa, ecc.) e che sono utilizzabili quando il risultato di una certa attività non è rappresentato da un prodotto tangibile;

relazioni tra il costrutto ed altri costrutti. Un tale tipo di studio può riguardare sia strumenti sviluppati, allo scopo di valutare una teoria esistente (la situazione migliore), che strumenti sviluppati, allo scopo di valutare un costrutto che non fa parte di una teoria esistente.

L'indicatore operativo del grado di corrispondenza tra lo strumento e il criterio è di solito rappresentato dalla dimensione della loro correlazione; in altre parole il confronto tra misura e criteri è realizzato e misurato in termini di analisi della correlazione, estesa anche al caso multivariato: il livello di correlazione ottenuto specifica il grado di validità; per questo motivo in tale contesto il coefficiente di correlazione viene indicato spesso come *coefficiente di validità*: se i risultati statistici conducono a decisioni e scelte ragionevoli allora il gruppo di variabili in questione ha validità predittiva<sup>2</sup>.

I complessi ambiti della ricerca sociale non consentono una stima accurata e, ragionevolmente, producono il più delle volte modeste correlazioni tra strumento e criterio esterno<sup>3</sup>.

La misura e il criterio esterno possono essere misurati in momenti diversi; in relazione a ciò è possibile distinguere tra:

- criterio *post-diction*: quando il criterio esterno viene misurato dopo l'applicazione dello strumento da validare;
- criterio *concomitante*: quando il criterio esterno viene misurato contemporaneamente allo strumento da validare;
- criterio *predittivo*: quando il criterio esterno viene misurato prima dell'applicazione dello strumento da validare.

Tale distinzione riguarda solo l'applicazione dello strumento, in quanto la logica e le procedure di validazione nei tre tipi non sono diverse: la validità predittiva è determinata solamente dal grado di corrispondenza tra i dati, indipendentemente da quando sono stati rilevati. E' importante riconoscere che l'utilità scientifica e pratica della validazione predittiva dipende sia dalla misurazione del criterio che dalla qualità dello strumento di misurazione stesso.

E' importante inoltre che gli studi di validità presentino i seguenti due aspetti:

- a. la teoria sulla quale sono basate le ipotesi deve essere solida;
- b. il criterio utilizzato deve essere misurato in modo affidabile.

Come si è visto la logica alla base di tale validità è abbastanza semplice e chiara; è però altrettanto chiaro che le procedure di validazione predittiva presentano una limitata utilità in quanto in molte situazioni non esiste alcun criterio con cui valutare in modo ragionevole la misura. Più è astratto il concetto, più difficile è trovare un criterio appropriato per la valutazione della misura.

La validità predittiva diviene particolarmente importante quando a livello pratico la misura da validare ha funzioni *operative* e *decisionali* (assunzione di personale, creazione di servizi sulla base di previste necessità di un gruppo di cittadini, ecc.).

#### **La validità di contenuto**

Per la validità di contenuto è essenziale che la misura da validare sia accettata come indicatore che

---

3. *subjective judgment criteria*, legati, a differenza delle precedenti tipologie oggettive, a giudizi di osservatori, le cui valutazioni soggettive possono essere tra loro correlate utilizzando metodi basati su confronti accoppiati, di ordinamento (*ranking* e *rating-scale*).

<sup>2</sup> Secondo un'altra strategia (*validità attraverso gruppi noti*), per tale validità si verifica l'ipotesi che determinati gruppi di soggetti ottengano punteggi più alti rispetto ad altri; la significatività della differenza tra i gruppi viene verificata calcolando la media dei punteggi ottenuti sullo strumento per ciascun gruppo e confrontando attraverso strumenti statistici le medie osservate (*t* di *Student*, analisi della varianza).

<sup>3</sup> Un esempio a riguardo può essere rappresentato da un test di maturità scolastica. Per poterlo completare occorre attendere la fine dell'anno. Relativamente a questo esempio occorre osservare che il profitto scolastico non dipende solo dalle capacità dello studente ma anche dalle caratteristiche degli insegnanti e della scuola: quindi anche per il criterio esterno si pongono problemi di validità e affidabilità. In questo caso si può procedere dimostrando la non-validità del criterio (confrontando i giudizi di più insegnanti sui risultati in una certa materia: se i giudizi divergono il criterio viene considerato non-valido).

definisce l'universo di contenuto di riferimento.

Tale validità viene verificata misurando il grado di rappresentatività degli indicatori che compongono la misura: se gli indicatori rappresentano un buon campione dell'universo degli indicatori dell'ambito da indagare, ovvero riflettono uno specifico dominio di contenuto, lo strumento ha una buona validità di contenuto.

La verifica della validità di contenuto è particolarmente importante e deve essere effettuata al momento della costruzione dello strumento e quindi prima dell'analisi vera e propria; questa dovrebbe fornire solo informazioni aggiuntive su tale validità, ovvero dovrebbe fornire solo un sostegno e una giustificazione statistica.

Per verificare la validità di contenuto occorre controllare e valutare i seguenti standard:

- Qualità e della rappresentatività degli indicatori: per valutare ciò è necessario identificare un profilo dettagliato delle dimensioni e dei concetti che devono essere rilevate. Il profilo deve essere collegato direttamente con l'ipotesi di ricerca che deve guidare e condurre alla formulazione e alla costruzione dello strumento. Tale valutazione è comunque complicata dal fatto che è difficile, se non impossibile, valutare il campionamento del contenuto.
- Adeguatezza e sensibilità della costruzione degli indicatori: per valutare ciò è possibile utilizzare alcune verifiche:
  - un alto livello di accordo (consistenza interna) tra gli indicatori dovrebbe rilevare che gli indicatori tendono a misurare qualcosa in comune; tale verifica però non garantisce la verifica della validità della costruzione: infatti gli indicatori che concorrono ad un unico contenuto possono anche misurare caratteristiche diverse non correlate;
  - il confronto tra diverse applicazioni dello strumento (per esempio prima e dopo una certa sperimentazione) consente di verificare se lo strumento misura l'evoluzione di un certo *fenomeno*; tale verifica richiede comunque un modello complesso e di difficile applicazione nel campo delle scienze sociali;
  - un alto livello di correlazione tra misura da validare e altri indicatori consente di concludere che lo strumento misura ciò che si vuole misuri; in realtà la registrazione e l'osservazione di alti livelli di correlazione non è garanzia di validità di contenuto: gli strumenti confrontati potrebbero misurare male nello stesso modo la stessa dimensione.

Come si è visto tali standard non possono essere sempre giudicati e valutati in modo adeguato; inoltre, nonostante i tentativi di applicare metodi statistici di verifica, la validità di contenuto è stabilita principalmente con altri criteri; la statistica può aiutare ma la validità di contenuto riguarda principalmente le caratteristiche riguardanti il contenuto e il modo con cui è presentato. E' per questo che è più corretto mettere a confronto i giudizi di diversi esperti in materia. In ogni caso, indipendentemente dai risultati dell'analisi, la decisione finale sulla validità di un indicatore o di un insieme di indicatori spetta al ricercatore.

#### ***La validità teorica o della struttura***

La validità della struttura è direttamente collegata alle teorie di riferimento per il ricercatore; in particolare essa riguarda il rapporto tra lo strumento e le dimensioni e i concetti da misurare e presuppone una solida base di preparazione, studio e sperimentazione sull'argomento. Da questo punto di vista un gruppo di variabili è valido se in esso trovano conferma le teorie di riferimento. Le teorie variano rispetto alla *specificità* e alla *grandezza* dell'area che gli indicatori devono ricoprire. Come sappiamo in alcuni casi il «dominio» è talmente piccolo che uno qualsiasi degli indicatori identificati è sufficiente a coprire l'area; maggiore è l'area da ricoprire, maggiore è la difficoltà nel definire quali indicatori appartengono all'area.

Lo strumento ritenuto valido nella struttura può essere utilizzato da coloro che accettano l'interpretazione teorica di ciò che esso rappresenta. Naturalmente il costrutto sottostante continua a rimanere un'entità teorica. La concezione della natura del costrutto e le ragioni delle relazioni con altri costrutti sono basate su una struttura teorica che può essere successivamente sostituita da una nuova struttura che ridefinisce il costrutto.

#### 4. Affidabilità e validità

La verifica della validità della struttura è un problema che non può essere risolto statisticamente ma logicamente; comunque, anche se la verifica della validità teorica è essenzialmente di tipo logico, questa è affiancata da una verifica statistica. Tale verifica mira a determinare quanto gli indicatori tendono a misurare la stessa cosa o cose diverse; ciò rappresenta una condizione necessaria ma non sufficiente per verificare la validità teorica.

Gli strumenti statistici utilizzati per studiare la validità teorica possono essere espressi in termini di

- *consistenza interna*, ovvero sulla base della tendenza di misure diverse a correlare molto tra loro e ad essere influenzate allo stesso modo da trattamenti sperimentali;
- determinazione della correlazione della misura del costrutto con altre misure relative ad altri costrutti;
- *analisi fattoriale*.

Il prodotto finale di tale processo dovrebbe condurre ad un costrutto

- a. ben definito in termini di osservazione,
- b. ben rappresentato in termini di variabili osservabili,
- c. eventualmente correlato con altri costrutti.

E' comunque molto importante tener presente che la validità di costrutto può essere sostenuta ma mai provata; è infatti possibile che successivi lavori e studi possano dare nuove interpretazioni dei risultati e verificare nuovi modelli.

Con il seguente schema cerchiamo di sintetizzare quali possono essere le finalità dei diversi approcci alla validità nel caso di indicatori soggettivi:

PROBLEMA	PROCEDIMENTO PER LA STIMA DELLA VALIDITA'	APPLICATO SOPRATTUTTO A
<b>VALIDITA' PREDITTIVA E CONCORRENTE</b>		
Come si correlano criterio e punteggio?	Somministrato lo strumento, dopo un certo periodo si raccolgono i dati relativi al criterio	<ol style="list-style-type: none"> <li>1. strumenti per la selezione di soggetti;</li> <li>2. strumenti usati in campo scolastico e clinico</li> <li>3. strumento usato per valutazioni complesse</li> </ol>
<b>VALIDITÀ DI CONTENUTO</b>		
Gli strumenti riescono ad esplorare e penetrare le situazioni che dovrebbero valutare?	Si analizzano gli item e le risposte richieste, in relazione all'oggetto che lo strumento dovrebbe valutare.	<ol style="list-style-type: none"> <li>1. strumenti usati per valutare i programmi educativi.</li> <li>2. Tecniche di osservazione per lo studio del comportamento abituale</li> </ol>
<b>VALIDITÀ DI COSTRUTTO</b>		
In che modo interpretare i punteggi ottenuti? Lo strumento misura proprio le variabili che dovrebbero misurare?	Si elaborano delle ipotesi sul significato da attribuire ai punteggi dello strumento, stabilendo il grado secondo cui i punteggi alti debbano differire da quelli bassi o quali variabili possono alterarli. Ciascuna di queste ipotesi viene verificata singolarmente.	<ol style="list-style-type: none"> <li>1. strumenti usati per valutare intelligenza o personalità sia a fini diagnostici e di orientamento individuale, che per ricerca e valutazione scolastica.</li> </ol>

#### Verifica contemporanea della validità convergente e discriminante

Come abbiamo visto, la *validità*

- *convergente* è quella verificata dall'osservazione di un'alta correlazione osservata tra la misura da validare e altre misure di costrutti teoricamente legati al primo, e
- *discriminante* è quella verificata dalla osservazione di una modesta o nulla correlazione tra la misura da validare e altre misure di altri costrutti non legati a quello misurato.

Questi due tipi di validità possono essere studiati in relazione tra loro, quando è possibile fare le due ipotesi contemporaneamente. L'idea di base è che un costrutto debba essere correlato maggiormente con se stesso che con altri.

Nella validità convergente ci si aspetta un'alta correlazione tra misure di costrutti teoricamente legati. Idealmente dovrebbero registrare circa lo stesso livello di affidabilità. Siccome si ipotizza sempre un certo livello di errore che abbassi il livello di affidabilità, tali correlazioni osservate non raggiungono mai il massimo livello.

A metà del Novecento Cronbach ed altri elaborarono, nell'ambito della definizione degli Standard per lo sviluppo di misure in psicologia per l'*Associazione Americana di Psicologia*, un particolare

approccio alla verifica della validità, la *rete nomologica*; tale rete di verifica dovrebbe comprendere

- a. la struttura teorica relativa a ciò che si cerca di misurare,
- b. la struttura empirica relativa al come si cerca di misurare,
- c. la specificazione dei legami tra le due precedenti strutture.

Tale rete è basata su un certo numero di principi che guida il ricercatore nella verifica; ciò che l'approccio cerca di fare è di collegare l'ambito concettuale/teoretico con quello osservativo. Se da una parte l'idea di rete nomologica può funzionare a livello filosofico, dall'altra essa non fornisce una metodologia pratica e utilizzabile per verificare realmente la validità.

Successivamente Campbell e Fiske nel 1959 hanno sviluppato un particolare approccio, definito *Multitrait-Multimethod Matrix (MTMM)*, che ha dato un impulso verso una particolare metodologia di verifica della validità. Il *MTMM* consente di indagare simultaneamente la validità convergente e la validità discriminante; l'applicazione del *MTMM* richiede, come vedremo nel paragrafo successivo, che vengano misurati almeno due costrutti e che ciascuno di essi sia misurato con almeno due metodi distinti.

## 4.2 VALIDITÀ, AFFIDABILITÀ E INDICATORI MULTIPLI

Secondo alcuni autori (Marradi, 1980) l'analisi della consistenza interna di un gruppo di indicatori è "una proprietà intermedia tra attendibilità e validità [...] e può essere misurata, dato che non fa riferimento a qualcosa di esterno alla matrice dei dati; può essere indizio di validità ma non certo una prova". Ciò ha indotto ad un certo scetticismo nei confronti di misure di affidabilità e validità basate sulla congruenza interna di batterie o altri gruppi di indicatori operativamente definiti in maniera troppo simile.

I primi a mettere in luce il fatto che nell'analisi della consistenza interna gli aspetti di affidabilità e di validità sono inestricabilmente legati sono stati Campbell e Fiske nel 1959. Seguiamo il loro ragionamento; essi si sono chiesti se la consistenza interna di un gruppo di indicatori fosse dovuta alla prossimità semantica o piuttosto solo alla somiglianza delle loro definizioni operative<sup>4</sup>. Per affrontare tale problema sarebbe necessario utilizzare tecniche di rilevazione più raffinate.

Relativamente ad un determinato concetto, si dovrebbero scegliere quegli indicatori che possano essere registrati con definizioni operative quanto più possibile diverse l'una dall'altra (fonte delle informazioni, forma delle domande, tipo di risposte precodificate, strumenti e situazioni in cui registrare le risposte, ecc.). Se le correlazioni tra gli indicatori dello stesso concetto registrati con definizioni operative diverse fossero considerevolmente più alte delle correlazioni tra gli indicatori di concetti diversi registrati con definizioni operative simili, il problema sarebbe risolto. Si realizzerebbe in questo modo una validazione *convergente* perché indicatori formalmente simili ma legati semanticamente a concetti diversi risulterebbero scarsamente congruenti tra loro. Quindi il requisito necessario per verificare la congruenza tra indicatori deve derivare dalla loro vicinanza semantica e non dalla somiglianza tra le relative definizioni operative.

Tale approccio mette in crisi la capacità delle scienze sociali di conoscere il proprio oggetto senza alterarlo; la definizione operative appare più che un modo per registrare gli stati degli oggetti su una certa proprietà, un meccanismo che contribuisce a creare gli stati che registra.

Tale approccio ha condotto a ad un generale adeguamento alla richiesta di dare definizioni operative quanto più differenti degli indicatori dello stesso concetto; tale adeguamento richiede spesso una certa creatività metodologica e pone problemi pratici non sempre superabili.

---

<sup>4</sup> Ciò si pone per esempio quando in questionario le domande sono proposte in successione dallo stesso intervistatore e l'intervistato deve scegliere per ciascuna domanda una risposta in un limitato arco di alternative pre-codificate, favorendo il presentarsi del *response-set*.

### 4.2.1 Il metodo proposto da Campbell e Fiske

Come si è già detto, quello della validità rappresenta il problema più critico nella ricerca empirica<sup>5</sup> e riguarda il problema di come gli indicatori misurino il concetto astratto derivato dalla teoria; tale problema non consente di avere sicurezze e rimane in un qualche modo discutibile e incerto. Secondo Campbell e Fiske è possibile confidare sulle caratteristiche delle misure di cui si dispone e su come esse correlano tra loro. Se si misura un certo tratto, o concetto astratto, con diverse metodologie tra loro molte diverse e se tali diverse procedure producono risultati che sono abbastanza simili, è possibile confidare nella validità delle misure, potendo affermare, in modo ragionevole e convincente, che ciascuno dei metodi produce una misura valida della caratteristica in questione. Con l'approccio definito da Campbell e Fiske i concetti di *affidabilità* e di *validità* trovano la seguente nuova definizione:

*L'affidabilità rappresenta l'accordo tra due tentativi di misurare lo stesso tratto attraverso metodi tra loro il più possibile simili. La validità è rappresentata dall'accordo tra due tentativi di misurare lo stesso tratto attraverso metodi tra loro il più possibile diversi.*

Tale definizione contiene le basi logiche della metodologia da loro proposta.

Nella teoria classica della misurazione, come abbiamo visto, si fa distinzione tra punteggi veri e punteggi osservati; sulla base degli assunti riguardanti tali punteggi è definito un coefficiente di affidabilità. Per poter calcolare i coefficienti basati sui metodi tradizionali è necessario, come sappiamo, poter disporre di almeno due misurazioni; in altre parole la valutazione dell'affidabilità di una misura è realizzabile attraverso diversi schemi sperimentali che alla base presentano la nozione di ripetibilità ovvero di applicazione di *metodi tra loro il più possibile simili*.

Con l'approccio proposto da Campbell e Fiske ci si sposta dalla nozione di pura affidabilità (attraverso metodi il più possibile simili tra loro) alla nozione di validità (metodi il più possibile diversi tra loro).

Per poter definire l'affidabilità e la validità, i due autori hanno ridefinito le nozioni di validità convergente e discriminante:

- metodi diversi di misurazione possono convergere verso la misurazione dello stesso tratto (*validità convergente*);
- stessi metodi di misurazione possono non correlare perché misurano tratti diversi (*validità discriminante*).

Il metodo da loro proposto può essere formalmente così presentato.

Il valore  $y_{ij}$  dell'indicatore  $i$ -esimo raccolto con il metodo  $j$ -esimo può essere decomposto in due componenti:

- una stabile ( $T_{ij}$ ), corrispondente al punteggio vero nella teoria classica,
- una casuale ( $e_{ij}$ ).

La risposta e le sue due componenti si legano tra loro nel modo seguente:

$$y_{ij} = h_{ij}T_{ij} + e_{ij}$$

dove

$h_{ij}$  livello di relazione tra componente stabile (punteggio vero) e risposta.

Il punteggio vero può essere ulteriormente decomposto in tre componenti:

- una che rappresenta il punteggio sulla variabile che interessa  $F_i$ ,
- una dovuta al metodo utilizzato  $M_j$ ,
- una dovuta alla combinazione di metodo e tratto  $u_{ij}$ ;

<sup>5</sup> In un certo senso tale problema è simile a quello dell'attribuzione di una etichetta a ciascun fattore nell'analisi fattoriale.

dopo la standardizzazione ciò conduce alla seguente equazione:

$$T_{ij} = b_{ij}F_i + g_{ij}M_j + u_{ij}$$

dove

$b_{ij}$  livello di relazione tra la variabile latente di interesse e il punteggio vero

$g_{ij}$  effetto della componente metodo sul punteggio vero

Sapendo che tutte le variabili, eccetto i termini di disturbo, sono standardizzate e che il metodo e i fattori non sono correlati, i coefficiente  $h_{ij}$ ,  $b_{ij}$  e  $g_{ij}$  indicano la forza delle relazioni tra le variabili nel modello; a tali coefficienti è stata data una speciale interpretazione:

- $h_{ij}$  è chiamato *coefficiente di affidabilità*; il quadrato di tale coefficiente rappresenta una stima dell'affidabilità (*test-retest* nel senso della teoria classica);
- $b_{ij}$  è chiamato *coefficiente di validità del punteggio vero* in quanto il quadrato di tale coefficiente rappresenta la varianza spiegata nel punteggio vero attribuita alla variabile cui siamo interessati;
- $g_{ij}$  è chiamato *effetto del metodo* in quanto il quadrato di tale coefficiente rappresenta la varianza la varianza spiegata nel punteggio vero attribuita al metodo usato;
- la varianza di  $u_{ij}$  più  $g_{ij}^2$  a volta è chiamata *invalidità* in quanto è la varianza spiegata nel punteggio vero che non è dovuta alla variabile di interesse.

#### 4.2.2 La matrice Multi-Trait Multi-Method

Per poter stimare l'errore di misurazione come è stato definito qui è necessario misurare almeno tre tratti con tre diversi metodi. Tale disegno, introdotto da Campbell e Fiske, è chiamato *Multi-Trait-Multi-Method (MTMM)* in quanto ciascun tratto di un numero di tratti (costrutti) è misurato con un numero di metodi differenti. Per analizzare i dati tratti dal disegno sperimentale *MTMM* è stato definito un modello causale. Oltre la distinzione tra punteggi veri e punteggi osservati, sono introdotti fattori latenti sia per il fattore "tratto" che per il fattore "metodo". Si assume che

- i fattori "tratto" siano tra loro correlati ( $\rho(F_1, F_2)$ ),
- i fattori "metodo" non siano correlati tra loro,
- i fattori "metodo" non siano correlati con i fattori "tratto".

Il requisito principale per poter applicare la matrice *multitrait-multimethod (MTMM)* è quello di poter disporre di almeno di tre differenti caratteristiche, ciascuna delle quali misurata con metodi tra loro molto diversi<sup>6</sup>. Per poter procedere occorre quindi innanzi tutto identificare i tratti e i metodi.

##### La costruzione della matrice

- Identificazione dei tratti: supponiamo di voler studiare le valutazioni individuali riguardo ai due partiti politici misurando i seguenti tratti:
  - a. valutazione del partito,
  - b. ideologia politica (per esempio *liberale-moderata-conservatrice*),
  - c. livello di coinvolgimento politico.
- Identificazione dei metodi: per misurare ciascuno dei tratti individuati è possibile utilizzare i medesimi tre metodi:
  1. *scala di atteggiamento* o *questionario*;
  2. *osservazione partecipata*, realizzata trascorrendo due o tre giorni con ciascun soggetto

<sup>6</sup> Come per altri approcci qui proposti, le caratteristiche da misurare possono essere rappresentate da atteggiamenti, comportamenti e possono riguardare sia individui che aggregazioni quali istituzioni, organizzazioni, città, nazioni, ecc.

#### 4. Affidabilità e validità

durante una campagna elettorale, registrando fedelmente qualsiasi accenno ad uno dei partiti; successivamente sarà possibile registrare il numero di affermazioni positive e negative date da ciascun soggetto e dirette a ciascun partito;

3. utilizzazione degli *informatori* che consiste nel chiedere ad amici e a parenti quali sono secondo loro le valutazioni riguardanti ciascun soggetto.

Tale quadro legittima l'utilizzo dell'approccio *MTMM*. Il primo passaggio dell'analisi finalizzata a stimare l'affidabilità, la validità e gli effetti dei metodi è quello di calcolare la matrice di correlazione<sup>7</sup> per le misure utilizzate. Tale matrice *MTMM* potrebbe essere rappresentata nel modo seguente:

		1			2			3		
METODO	TRATTI	A	B	C	A	B	C	A	B	C
1	A	$r_{A_1A_1}$								
	B	$r_{B_1A_1}$	$r_{B_1B_1}$							
	C	$r_{C_1A_1}$	$r_{C_1B_1}$	$r_{C_1C_1}$						
2	A	$r_{A_2A_1}$	$r_{A_2B_1}$	$r_{A_2C_1}$	$r_{A_2A_2}$					
	B	$r_{B_2A_1}$	$r_{B_2B_1}$	$r_{B_2C_1}$	$r_{B_2A_2}$	$r_{B_2B_2}$				
	C	$r_{C_2A_1}$	$r_{C_2B_1}$	$r_{C_2C_1}$	$r_{C_2A_2}$	$r_{C_2B_2}$	$r_{C_2C_2}$			
3	A	$r_{A_3A_1}$	$r_{A_3B_1}$	$r_{A_3C_1}$	$r_{A_3A_2}$	$r_{A_3B_2}$	$r_{A_3C_2}$	$r_{A_3A_3}$		
	B	$r_{B_3A_1}$	$r_{B_3B_1}$	$r_{B_3C_1}$	$r_{B_3A_2}$	$r_{B_3B_2}$	$r_{B_3C_2}$	$r_{B_3A_3}$	$r_{B_3B_3}$	
	C	$r_{C_3A_1}$	$r_{C_3B_1}$	$r_{C_3C_1}$	$r_{C_3A_2}$	$r_{C_3B_2}$	$r_{C_3C_2}$	$r_{C_3A_3}$	$r_{C_3B_3}$	$r_{C_3C_3}$

I valori della matrice rappresentano coefficienti di correlazione; l'analisi di tale matrice richiede, a causa del diverso significato che le correlazioni assumono, la suddivisione delle correlazioni in quattro diversi raggruppamenti:

- Diagonale di affidabilità: same-trait same-method**, composto dalle correlazioni tra due tentativi di misurare la stessa caratteristica utilizzando lo stesso metodo; all'interno della matrice esse corrispondono ai valori  $r_{A_1A_1}$ ,  $r_{B_1B_1}$ ,  $r_{C_1C_1}$ ,  $r_{A_2A_2}$ ,  $r_{B_2B_2}$ ,  $r_{C_2C_2}$ ,  $r_{A_3A_3}$ ,  $r_{B_3B_3}$ ,  $r_{C_3C_3}$  (valori su sfondo blu); tali correlazioni possono essere interpretate come coefficienti di affidabilità<sup>8</sup>.
- Diagonali di validità: same-trait different-method**, composto dalle correlazioni tra la stessa caratteristica misurata con metodi diversi; all'interno della matrice esse corrispondono ai valori di  $r_{A_2A_1}$ ,  $r_{B_2B_1}$ ,  $r_{C_2C_1}$ ,  $r_{A_3A_1}$ ,  $r_{A_3A_2}$ ,  $r_{B_3B_1}$ ,  $r_{B_3B_2}$ ,  $r_{C_3C_1}$ ,  $r_{C_3C_2}$  (valori in diagonale su sfondo verde) e che possono essere interpretati come coefficienti di validità convergente in quanto rappresentano le correlazioni tra le stesse caratteristiche misurate attraverso metodi tra loro molto diversi, coerentemente con la concettualizzazione di validità fatta da Campbell e Frisse: se si misura una particolare caratteristica, correttamente concettualizzata, in modi diversi e si ottengono gli stessi risultati, le procedure di misurazione risultano probabilmente valide. Tali

<sup>7</sup> Il tipo di correlazione scelta può influenzare i risultati; nel caso di variabili categoriche si suggerisce di utilizzare, per evitare grandi *bias*, le correlazioni policoriche e poliseriali.

<sup>8</sup> Tali valori possono essere prodotti da metodi quali il *test-retest* o lo *split-half* (v. *I modelli di scaling*, n. 2 della serie AstRis).

valori dovrebbero essere statisticamente rilevanti e di grande dimensione. Per ciascun metodo il valore della validità dovrebbe essere maggiore di qualsiasi altro valore nella riga e nella colonna corrispondenti; in altre parole due misure dello stesso costrutto dovrebbero avere correlazioni più strette tra loro che con le misure di altri costrutti.

- C. *Triangoli Different-trait same-method*, composto dalle correlazioni tra diverse caratteristiche misurate con lo stesso metodo; all'interno della matrice esse corrispondono ai valori di  $r_{B_1A_1}$ ,  $r_{C_1A_1}$ ,  $r_{C_1B_1}$ ,  $r_{B_2A_2}$ ,  $r_{C_2A_2}$ ,  $r_{C_2B_2}$ ,  $r_{B_3A_3}$ ,  $r_{C_3A_3}$ ,  $r_{C_3B_3}$  (triangolo con sfondo rosso). Campbell e Friske chiamano tali triangoli *heterotrait-monomethod*. L'ordine di grandezza di tali correlazione (all'interno dei triangoli) dovrebbe essere lo stesso.
- D. *Triangoli Different-trait different-method*, composto dalle correlazioni tra caratteristiche diverse misurate con metodi differenti; all'interno della matrice esse corrispondono ai valori  $r_{A_2B_1}$ ,  $r_{A_2C_1}$ ,  $r_{B_2A_1}$ ,  $r_{B_2C_1}$ ,  $r_{C_2A_1}$ ,  $r_{C_2B_1}$ ,  $r_{A_3B_1}$ ,  $r_{A_3C_1}$ ,  $r_{A_3B_2}$ ,  $r_{A_3C_2}$ ,  $r_{B_3A_1}$ ,  $r_{B_3C_1}$ ,  $r_{B_3A_2}$ ,  $r_{B_3C_2}$ ,  $r_{C_3A_1}$ ,  $r_{C_3B_1}$ ,  $r_{C_3A_2}$ ,  $r_{C_3B_2}$  (sfondo fucsia).

Ricapitolando, le categorie possono essere visualizzate nella matrice nel modo seguente:

- same-trait same-method* (sfondo blu);
- same-trait different-method* (sfondo verde);
- different-trait same-method* (sfondo rosso);
- different-trait different-method*: (sfondo fucsia).

#### *L'interpretazione della matrice: i criteri*

Per poter interpretare i valori all'interno della matrice *MTMM*, Campbell e Friske hanno individuato quattro criteri:

- I coefficienti di affidabilità dovrebbero rappresentare i valori più alti della matrice.*
- I coefficienti di validità dovrebbero essere significativamente diversi da 0 e sufficientemente grandi da incoraggiare ulteriori analisi di validità.*

Nella matrice dell'esempio i valori di tali coefficienti vanno da .39 a .68; applicando un semplice test di significatività a tali correlazione dovrebbe essere possibile accertarsi del livello di significatività loro associato. Una volta osservato un livello soddisfacente di significatività sarà possibile esplorare gli altri criteri.

- Ciascun coefficiente di validità dovrebbe essere maggiore di tutte le correlazioni different-trait different-method presenti nella stessa riga o nella stessa colonna del coefficiente di validità.*

Nella matrice dell'esempio il valore del coefficiente di validità per *A1-A2* è .51. Il confronto rilevante è quindi tra .51 e i coefficienti che sono nella stessa riga o colonna del valore .51 nei due triangoli adiacenti. Tali quattro coefficienti sono .32, .29, .31 e .30, tutti più piccoli del coefficienti di validità rilevante.

- Ciascun coefficiente di validità dovrebbe essere maggiore del corrispondente coefficiente different-trait same-method.*

La motivazione di tale criterio è data dal fatto che, perché delle misure siano valide vi deve essere più varianza di tratto che varianza di metodo: se le caratteristiche sono tra loro veramente distinte a livello concettuale allora la maggior parte della varianza condivisa dovrebbe riflettere la varianza metodologica<sup>9</sup>. La stessa caratteristica misurata con metodi diversi dovrebbe riflettere principalmente la varianza di tratto che dovrebbe essere maggiore della varianza di metodi delle correlazioni *different trait-same method*.

Nella matrice dell'esempio il valore del coefficiente di validità per la caratteristica *A* ovvero *A1-*

<sup>9</sup> Strettamente parlando, le correlazioni *different trait-same method* dovrebbero comprendere più della varianza metodologica. Anche se i tratti possono essere concettualmente distinti, vi può essere tra loro probabilmente qualche relazione causale. In genere, si dovrebbe utilizzare l'approccio *M-M* con variabili che siano entrambe concettualmente distinte e che hanno tra loro solamente connessioni causali minori e non dirette. Se ciò risultasse vero, la varianza comune tra lo stesso concetto misurato attraverso diversi metodi dovrebbe essere maggiore della varianza comune tra concetti diversi misurati attraverso lo stesso metodo.

$A_2$  è uguale a .51; tale valore dovrebbe essere maggiore delle correlazioni che coinvolgono  $A_1$  o  $A_2$  con altre caratteristiche ma con lo stesso metodo. Quindi .51 viene confrontato con le correlazioni tra  $A_1$  e  $B_1$ ,  $A_1$  e  $C_1$ ,  $A_2$  e  $B_2$  e  $A_2$  e  $C_2$ , rispettivamente .42, .38, .44 e .38: il criterio può dirsi soddisfatto.

- E. Si dovrebbe osservare lo stesso modello di correlazioni all'interno di ciascun triangolo,
- sia quelli con gli elementi che presentano correlazioni tra tratti diversi che utilizzano metodi diversi,
  - sia quelli che riflettono correlazioni tra tratti diversi utilizzando lo stesso metodo.

Nella matrice dell'esempio nel triangolo in alto (che chiameremo triangolo 1-1) il modello, in ordine inverso di grandezza di correlazione, è

$$A_1-B_1 (.42), A_1-C_1 (.38) \text{ e } B_1-C_1 (.33).$$

Tale modello di correlazioni dovrebbe riflettere le varie connessioni causali tra i tratti  $A$ ,  $B$  e  $C$ . Tutte le correlazioni dovrebbero riflettere una parte della varianza comune di metodi, in tale triangolo, ma il modello di ineguaglianza dovrebbe riflettere anche la forza delle connessioni causali tra i tratti. Secondo il modello del triangolo 1-1 la correlazione tra  $A$  e  $B$  dovrebbero essere maggiore che non tra  $A$  e  $C$  o tra  $B$  e  $C$ .

Vediamo un esempio<sup>10</sup>:

		1			2			3		
METODO	TRATTI	A	B	C	A	B	C	A	B	C
1	A	(.082)								
	B	0.42	(.079)							
	C	0.38	0.33	(.074)						
2	A	0.51	0.32	0.29	(.069)					
	B	0.31	0.45	0.19	0.44	(.084)				
	C	.030	0.25	0.39	0.38	0.32	(.065)			
3	A	0.58	0.31	0.30	0.62	0.36	0.28	(.089)		
	B	0.35	0.48	0.21	0.25	0.68	0.25	0.46	(.075)	
	C	0.28	0.19	0.39	0.24	0.23	0.59	0.37	0.36	(.068)

- A. La correlazione più bassa nella diagonale di validità è statisticamente significativa a livello 0.05.
- B. I valori di tali coefficienti vanno da 0.39 a 0.68; applicando un semplice test di significatività a tali correlazione dovrebbe essere possibile accertarsi del livello di significatività loro associato. Una volta osservato un livello soddisfacente di significatività sarà possibile esplorare gli altri criteri.
- C. Il valore del coefficiente di validità  $r_{A_1A_2}$  è 0.51. Il confronto rilevante è quindi tra 0.51 e i coefficienti che sono nella stessa riga o colonna del valore 0.51 nei due triangoli adiacenti. Tali quattro coefficienti sono 0.32, 0.29, 0.31 e 0.30, tutti più piccoli del coefficienti di validità rilevante.
- D. Il valore del coefficiente di validità per la caratteristica  $A$  ( $r_{A_1A_2}$ ) è uguale a 0.51; tale valore dovrebbe essere maggiore delle correlazioni che coinvolgono per la caratteristica  $A$  i metodi 1 e 2 con altre caratteristiche ma con lo stesso metodo. Quindi 0.51 viene confrontato con le correlazioni  $r_{A_1B_1}$ ,  $r_{A_1C_1}$ ,  $r_{A_2B_2}$  e  $r_{A_2C_2}$ , rispettivamente 0.42, 0.38, 0.44 e 0.38: il criterio può dirsi soddisfatto.
- E. Nella matrice dell'esempio nel triangolo in alto (che chiameremo triangolo 1-1) il modello, in

<sup>10</sup> I dati nella tabella non corrispondono ai dati ottenuti a partire da uno studio reale e sono presentati solo a titolo illustrativo.

ordine inverso di grandezza di correlazione, è

$$r_{A_1B_1} (.42), r_{A_1C_1} (.38) \text{ e } r_{B_1C_1} (.33).$$

Tale modello di correlazioni dovrebbe riflettere le varie connessioni causali tra i tratti *A*, *B* e *C*. Tutte le correlazioni dovrebbero riflettere una parte della varianza comune di metodi, in tale triangolo, ma il modello di ineguaglianza dovrebbe riflettere anche la forza delle connessioni causali tra i tratti. Secondo il modello del triangolo 1-1 la correlazione tra *A* e *B* dovrebbero essere maggiore che non tra *A* e *C* o tra *B* e *C*.

Si può dire che i dati presentati nell'esempio soddisfano tutti i criteri.

In particolare, i confronti da effettuare per verificare i **criteri B e C** sono presentati nella seguente tabella; essendo il valore del coefficiente di validità sempre maggiore di tutti i confronti rilevanti *different-trait, different-method* e *different-trait, same-method*, i criteri **B e C** sono stati soddisfatti in tutti i casi. All'interno di tutti i triangoli presenti nella tabella iniziale il modello è esattamente lo stesso: *A* e *B* evidenziano le correlazioni maggiori, successivamente *A* e *C* mentre *B* e *C* presentano le correlazioni più basse. Essendo ciò vero indipendentemente dalla combinazione dei metodi coinvolti nel confronto, è stato verificato anche il **criterio D**.

A. Criteri B e C		Coefficienti di validità	Confronti per il criterio B				Confronti per il criterio C			
	$r_{A_2A_1}$	0.51	0.32	0.29	0.31	0.30	0.42	0.38	0.44	0.38
	$r_{B_2B_1}$	0.45	0.32	0.19	0.31	0.25	0.42	0.33	0.44	0.32
	$r_{C_2C_1}$	0.39	0.29	0.19	0.30	0.25	0.38	0.33	0.38	0.32
	$r_{A_3A_2}$	0.62	0.36	0.28	0.25	0.24	0.44	0.38	0.46	0.37
	$r_{B_3B_2}$	0.68	0.36	0.25	0.25	0.23	0.44	0.32	0.46	0.36
	$r_{C_3C_2}$	0.59	0.28	0.25	0.24	0.23	0.38	0.32	0.37	0.36
	$r_{A_3A_1}$	0.58	0.31	0.30	0.35	0.28	0.42	0.38	0.46	0.37
	$r_{B_3B_1}$	0.48	0.31	0.21	0.35	0.19	0.42	0.33	0.46	0.36
	$r_{C_3C_1}$	0.39	0.30	0.21	0.28	0.19	0.38	0.33	0.34	0.36

B. Criterio D	Triangolo	Ordine
	1-1	$r_{B_1A_1} \Rightarrow r_{C_1A_1} \Rightarrow r_{C_1B_1}$
	2-2	$r_{B_2A_2} \Rightarrow r_{C_2A_2} \Rightarrow r_{C_2B_2}$
	3-3	$r_{B_3A_3} \Rightarrow r_{C_3A_3} \Rightarrow r_{C_3B_3}$
	1-2 top	$r_{A_2B_1} \Rightarrow r_{A_2C_1} \Rightarrow r_{B_2C_1}$
	1-2 bottom	$r_{B_2A_1} \Rightarrow r_{C_2A_1} \Rightarrow r_{C_2B_1}$
	1-3 top	$r_{A_3B_1} \Rightarrow r_{A_3C_1} \Rightarrow r_{B_3C_1}$
	1-3 bottom	$r_{B_3A_1} \Rightarrow r_{C_3A_1} \Rightarrow r_{C_3B_1}$
	2-3 top	$r_{A_3B_2} \Rightarrow r_{A_3C_2} \Rightarrow r_{B_3C_2}$
	2-3 bottom	$r_{B_3A_2} \Rightarrow r_{C_3A_2} \Rightarrow r_{C_3B_2}$

A conclusione di tale analisi, e sulla base di tali risultati, è possibile affermare che le procedure di misurazione utilizzate sono molto probabilmente valide e che misurano probabilmente ciò che si è

cercato di misurare.

Naturalmente in molte situazioni empiriche non tutte le verifiche effettuate e presentate nell'ultima tabella sono soddisfatte dai dati, pur disponendo di misure valide. Ciò può dipendere da molti fattori quali l'osservazione di livelli diversi di affidabilità e validità a causa di fluttuazioni casuali nel campionamento degli indicatori e dei soggetti.

Campbell e Fiske non hanno però definito dei parametri di riferimento su cui basarsi per stabilire se i dati si avvicinano ai criteri proposti.

Secondo alcuni tale modello può essere applicato anche al caso in cui i metodi non sono molto differenti tra loro come nel seguente esempio in cui si studia la validità di una nuova misura del livello di soddisfazione del lavoro confrontandola con uno standard esistente. Il *Job Satisfaction Survey* è uno strumento suddiviso in nove sottoscale ideate per valutare individualmente il livello di soddisfazione e gli atteggiamenti verso vari aspetti del lavoro.

È stato sviluppato specificamente per essere impiegato su lavoratori dipendenti nelle organizzazioni di servizi (ospedali, centri di igiene mentale, servizi sociali, ecc.). Nell'ambito della misurazione della soddisfazione del lavoro esiste una scala ben validata e largamente utilizzata chiamata *Job Descriptive Index (JDI)*; cinque delle sottoscale di *JSS* (metodo 2) sono comprese anche nella *JDI* (metodo 1). Dopo aver somministrato entrambe le scale ad un centinaio di soggetti, e disponendo di dati rilevati con due scale considerate metodi distinti, è stato possibile applicare il modello di analisi *MTMM*.

Nella seguente matrice *MTMM* sono presentati i risultati ottenuti per tre sottoscale: soddisfazione rispetto a natura dei compiti (*A*), retribuzione (*B*) e supervisione (*C*). La matrice contiene le correlazioni tra i 6 punteggi.

		METODO ⇒		JDI			JSS		
METODO ↓		TRATTI ↓		A	B	C	A	B	C
JDI	A	( )							
	B	.27	( )						
	C	.31	.23	( )					
JSS	A	.66	.24	.24	( )				
	B	.33	.62	.34	.29	( )			
	C	.25	.27	.80	.22	.34	( )		

I valori che risultano nell'esempio rivelano buone validità convergenti e discriminanti per la *JSS*. I valori di validità in diagonale sono tutti piuttosto alti (da .62 a .80) e sono i maggiori valori dell'intera matrice.

I valori delle corrispondenti correlazioni all'interno di ciascun triangolo sono abbastanza simili e sono piuttosto modesti e vanno da .22 a .34. Tali valori piuttosto bassi suggeriscono che le sottoscale valutano costrutti diversi.

### 4.2.3 Un esempio di applicazione del metodo

Anche se secondo molti ricercatori l'*achievement* (rendimento, realizzazione) rappresenta un concetto unidimensionale, alcuni hanno sottoposto tale assunto a verifica attraverso la matrice *MTMM*.

Tale applicazione non è molto ortodossa in quanto non sono stati selezionati tratti che a priori si ipotizzavano non correlati (o solamente molto correlati) ma si è cercato di specificare molti tipi diversi di un unico tratto (*achievement motivation*), che possono o non possono essere molto correlati. Sono stati definiti sei tipi diversi di *achievement* e sono state individuate cinque diverse strategie di misurazione. A scopo illustrativo qui ne sono presentate solo tre. I tre tipi di *achievement* selezionati:

1. *status with expert*, impegno e accettazione di perizie e giudizi di esperti;
2. *tendenza ad accumulare*, che comprende ricompense materiali come motivazione primaria all'*achievement*;
3. *interesse per la perfezione*, che comprende alti standard di *achievement* intellettuale e culturale.

I tre metodi selezionati sono:

- A. *self-rating*, in cui ciascun soggetto doveva valutare se stesso rispetto ad una serie di aggettivi polarizzati (tre per ogni tratto);
- B. *simulation role selection*: dopo avere descritto dei ruoli, uno per ciascun tratto, veniva chiesto ai soggetti, di selezionarne uno per la simulazione; il titolo e la descrizione dei ruoli sono stati fatti in modo accurato in modo che ciascuno riflettesse solamente uno dei tratti;
- C. *personality inventory* che comprendeva 38 item per ciascun tratto.

Nella seguente tabella sono presentate le correlazioni tra tratti e metodi:

		1			2			3		
METODO	TRATTI	A	B	C	A	B	C	A	B	C
1	A	(1.0)								
	B	-.16	(1.0)							
	C	.42	.07	(1.0)						
2	A	.30	-.09	.14	(1.0)					
	B	-.05	.34	.28	-.06	(1.0)				
	C	.15	-.19	.23	.19	-.04	(1.0)			
3	A	.52	.08	.33	.22	.12	.09	(1.0)		
	B	-.06	.69	.13	-.10	.38	.20	.17	(1.0)	
	C	.40	.02	.67	.11	.27	.18	.35	.22	(1.0)

I metodi sono piuttosto diversi, ma non è possibile stabilire se i tratti siano abbastanza differenti da validare le procedure di misurazione. Infatti non è stata utilizzata la matrice *MTMM* per validare le misure ma piuttosto per verificare l'unidimensionalità della motivazione all'*achievement*. Lo studio si conclude con una risposta negativa. Vediamo la tabella riguardante la verifica dei criteri 2, 3 e 4.

		Coefficienti di validità	Confronti per il criterio B			Confronti per il criterio C				
A. Criteri B e C	$r_{A_2A_1}$	.30	-.09	.14	-.05	.15	-.16	.42*	-.06	.19
	$r_{B_2B_1}$	.34	-.09	-.19	-.05	.28	-.16	.07	-.06	-.04
	$r_{C_2C_1}$	.23	.15	-.19	.14	.28*	.42*	.07	.19	-.04
	$r_{A_3A_2}$	.22	.12	.09	-.10	.11	-.06	.09	.17	.35*
	$r_{B_3B_2}$	.38	.12	.27	-.10	-.20*	-.06	-.04	.17	.22
	$r_{C_3C_2}$	.18	.11	.27*	.09	-.20*	.19*	-.04	.35*	.22*
	$r_{A_3A_1}$	.52	.08	.33	-.06	.40	-.16	.42	.17	.35
	$r_{B_3B_1}$	.69	.08	.02	-.06	.13	-.16	.07	.17	.22
	$r_{C_3C_1}$	.67	.40	.02	.33	.13	.42	.07	.35	.22

#### 4. Affidabilità e validità

		Triangolo	Ordine
		B. Criterio D	1-1
2-2			$r_{C_2A_2} \Rightarrow r_{C_2B_2} \Rightarrow r_{B_2A_2}$
3-3			$r_{C_3A_3} \Rightarrow r_{C_3B_3} \Rightarrow r_{B_3A_3}$
1-2	top		$r_{B_2C_1} \Rightarrow r_{A_2C_1} \Rightarrow r_{A_2B_1}$ **
1-2	bottom		$r_{C_2A_1} \Rightarrow r_{B_2A_1} \Rightarrow r_{C_2B_1}$ **
1-3	top		$r_{A_3C_1} \Rightarrow r_{B_3C_1} \Rightarrow r_{A_3B_1}$
1-3	bottom		$r_{C_3A_1} \Rightarrow r_{C_3B_1} \Rightarrow r_{B_3A_1}$
2-3	top		$r_{A_3B_2} \Rightarrow r_{A_3C_2} \Rightarrow r_{B_3C_2}$ **
2-3	bottom		$r_{C_3B_2} \Rightarrow r_{C_3A_2} \Rightarrow r_{B_3A_2}$ **

I valori registrati dai coefficienti di validità, che in tale tabella corrispondono alle correlazioni tra *same-trait different-method*, vanno da .18 a .69. Qui, con  $N=155$ , essi sono risultati tutti statisticamente significativi, anche se notiamo come il secondo metodo (selezione del ruolo di simulazione) risulta avere i coefficienti di validità più bassi degli altri due metodi. Ciò è dovuto al fatto che si tratta del metodo che più si differenzia dagli altri, in quanto mentre il *self-rating* e il *personality inventory* rappresentano delle auto-descrizioni, la selezione del ruolo è in un certo qual modo più di una misura di comportamento.

Il secondo criterio<sup>11</sup> è soddisfatto in modo soddisfacente in quanto ben 33 dei 36 confronti sono rientrati nel criterio con solamente tre correlazioni *different-trait different-method* più grandi dei loro corrispondenti coefficienti di validità; come si può notare le tre eccezioni coinvolgono il metodo 2.

Il terzo criterio<sup>12</sup> è soddisfatto solo in 31 dei 36 confronti; le cinque violazioni hanno coinvolto il metodo 2, particolarmente con i metodi 2 e 3 che misurano il tratto C.

Quindi due delle tre violazioni del secondo criterio e tre delle sei violazioni del terzo criterio coinvolgono il metodo 2 che può essere legittimamente sospettato di non essere valido in particolare quando è riferito al tratto C.

Relativamente al quarto criterio<sup>13</sup> notiamo come per cinque dei nove triangoli, i tratti A e C correlano molto tra loro, poi i tratti B e C, e i tratti A e B correlano molto debolmente.

Le inversioni suggeriscono l'esistenza di una qualche interazione tra metodi e tratti rivelando che alcuni metodi possano essere migliori misure di alcuni tratti.

Con tali risultati, i ricercatori possono concludere che, dati gli errori di misurazione e di campionamento, le misure sono generalmente valide e che possono essere utilizzate per ulteriori analisi.

Il metodo messo a punto da Campbell e Fiske per la valutazione della matrice funziona abbastanza bene in circostanze come quelle presentate, in cui la validità per tutti i metodi è abbastanza netta. In

<sup>11</sup> Ciascun coefficiente di validità dovrebbe essere maggiore di tutte le correlazioni *different-trait different-method* che sono nella stessa riga o colonna del coefficiente di validità, nei triangoli composti dai valori in corsivo.

<sup>12</sup> Ciascun coefficiente di validità dovrebbe essere maggiore delle correlazioni *different-trait same-method* che coinvolgono la stessa variabile come coefficiente di validità.

<sup>13</sup> Lo stesso modello di correlazioni dovrebbe essere evidenziato all'interno di ciascuno dei triangoli.

altri casi le deviazioni dal caso ideale lo rendono piuttosto difficile da applicare e interpretare. Sono disponibili pochi metodi statistici per la valutazione delle matrici *MTMM*. Quello diventato recentemente più popolare è quello che utilizza il modello delle equazioni strutturali. Tale tecnica può essere abbastanza utile ma presenta alcuni seri limiti. Il recente sviluppo di modelli *direct product* si presentano promettenti, sebbene si osservino poche applicazioni. Dovendo sviluppare misure, comunque, l'analisi della matrice di per sé può dare delle interessanti tracce sull'esistenza o meno di problemi di validità potenziale in alcune scale.

Per esplorare l'utilità di tale approccio in maggiore dettaglio, considerando il modello sottostante la matrice *MTMM* è possibile utilizzare la tecnica degli indicatori multipli per

- rappresentare le diverse misure e le variabili astratte;
- esaminare la struttura delle relazioni così prodotte

ovvero è possibile applicare i principi della *path analysis*. Ciò consente di esaminare simultaneamente gli assunti sottostanti e di illustrare l'estensione dei modelli con indicatori multipli a situazioni più complesse.

Mentre Campbell e Fiske presentavano l'utilizzo degli indicatori multipli per valutare la validità, Costner nel 1969 presentava un approccio alla valutazione dell'affidabilità che utilizza gli indicatori multipli, ponendo l'attenzione sulla questione dell'errore di misurazione.

#### ***4.2.4 Vantaggi e svantaggi dell'approccio MTMM***

Indubbiamente tale approccio fornisce una metodologia operativa per la valutazione della validità. In un'unica matrice è praticamente possibile esaminare simultaneamente sia la validità convergente che la validità discriminante. Nel cercare di includere metodi e tratti su un unico piano, i due ideatori hanno accentuato l'importanza della ricerca degli effetti del metodo di misurazione in aggiunta al contenuto della misurazione. In questo senso il *MTMM* fornisce una struttura rigorosa per valutare la validità.

Per molte ragioni, però, nonostante tali vantaggi, il *multitrait-multimethod* non ha trovato grandi applicazioni. Innanzitutto è necessario disporre di un disegno sperimentale di ricerca piuttosto complesso ed esteso (misurazione di più tratti attraverso diverse metodologie) non sempre pienamente realizzabile. Le decisioni finali, riguardanti la validità, sono sempre legate ad un giudizio individuale del ricercatore: ciò non sempre è apprezzato da chi ritiene che tale decisione sia presa "oggettivamente" per mezzo di un unico coefficiente statistico.