

Sapienza – Università di Roma



Dipartimento di Comunicazione e Ricerca Sociale

Dottorato di ricerca in Comunicazione Tecnologie Società

Curriculum in

Metodologia delle Scienze Sociali

XXVIII Ciclo

**La Coesione Sociale: un approccio integrato attraverso l'analisi
multilivello**

Gianmaria Bottoni

Tutor: Prof.ssa Isabella Mingo

INDICE

Introduzione.....	3
 1. Il concetto di coesione sociale	
1.1 La coesione sociale: un'introduzione teorica.....	5
1.2 Diverse definizioni del concetto di coesione sociale.....	6
1.2.1 L'approccio istituzionale.....	7
1.2.2 L'approccio policy-oriented.....	12
1.2.3 L'approccio accademico.....	16
1.3 Cos'è la coesione sociale? Una proposta.....	19
1.3.1 Relazioni fra individui.....	20
1.3.2 Relazioni individuo-gruppo.....	21
1.3.3 Relazioni individuo-stato.....	22
1.3.4 Lo schema della coesione sociale.....	22
 2. L'analisi multilivello e i modelli ad equazioni strutturali multilivello	
2.1 Introduzione.....	25
2.2 L'analisi multilivello.....	25
2.2.1 I presupposti dell'analisi multilivello.....	30
2.2.2 Il modello di regressione multilivello.....	34
2.2.3 Il modello nullo.....	36
2.2.4 Il modello ad intercetta casuale e con coefficiente di regressione casuale.....	37
2.2.5 La complessità delle equazioni di regressione multilivello.....	40
2.3 I modelli ad equazioni strutturali.....	43
2.3.1 Il modello strutturale e il modello di misurazione.....	44
2.3.2 Analisi fattoriale e analisi causale.....	47
2.3.3 Il modello completo.....	50
2.3.4 Le fasi dei modelli ad equazioni strutturali.....	51
2.4 Modelli multilivello ad equazioni strutturali.....	52
2.4.1 Perché applicare un approccio multilivello.....	52
2.4.2 Pregi e limitazioni dei modelli di regressione multilivello e dei modelli ad equazioni strutturali.....	53
2.4.3 La continuità fra i modelli di regressione multilivello e i modelli ad equazioni strutturali.....	54
2.4.4 I vantaggi dei modelli ad equazioni strutturali multilivello.....	55
2.4.5 La logica sottostante i modelli ad equazioni strutturali multilivello: la scomposizione della varianza e i diversi metodi di stima.....	58

2.4.6 Il modello di analisi fattoriale confermativa multilivello.....	62
2.4.7 Il modello MSEM completo: modello di misurazione e modello strutturale.....	63
2.4.8 Le strategie di analisi.....	65
3. Operativizzazione e misurazione del concetto di coesione sociale: un modello ad equazioni strutturali multilivello	
3.1 Una proposta di definizione operativa della coesione sociale.....	68
3.1.1 Considerazioni preliminari.....	68
3.1.2 La definizione operativa della coesione sociale.....	69
3.1.3 Dati, obiettivi e procedure.....	72
3.1.4 L'analisi fattoriale confermativa del modello di coesione sociale.....	74
3.1.5 Analisi fattoriale confermativa Multilivello.....	76
3.1.6 Modello 1: CFA multilivello.....	77
3.1.7 Modello 2: Ipotesi di invarianza cross-level con il vincolo dell'eguaglianza dei factor loadings.....	81
3.1.8 Modello 3: Ipotesi di invarianza di misurazione del modello di coesione sociale.....	83
3.1.9 Analisi fattoriale confermativa multilivello di secondo ordine.....	85
3.2 Le implicazioni del modello multilivello della coesione sociale.....	88
4. Le dimensioni della Coesione Sociale, la frammentazione etnica e la disuguaglianza: un modello ad equazioni strutturali multilivello.....	93
Conclusioni.....	99
Appendice statistica.....	101
Bibliografia.....	113

Introduzione

Il lavoro di tesi si snoda lungo l'arco di quattro capitoli. Nella prima parte dell'indagine, dopo un'attenta ricognizione della letteratura che mette in luce l'eterogeneità degli approcci teorici presenti in letteratura, viene presentata la mia prospettiva teorica generale impiegata per lo studio della coesione sociale. Accanto a tali riflessioni teoriche, viene fornito uno schema concettuale che, oltre ad indirizzare la presente indagine, si pone l'obiettivo di coadiuvare futuri ricercatori nelle operazioni di rilevazione e misurazione della coesione e, più generalmente, di fenomeni complessi. Lo schema suggerisce di analizzare la coesione in base ad una prospettiva che prenda in considerazione diverse dimensioni dell'analisi – micro, meso e macro – due componenti – soggettiva ed oggettiva – e i livelli a cui il concetto viene analizzato – individuale, di gruppo, aggregato. Lo schema si pone come base comune su cui futuri ricercatori potranno progressivamente sviluppare la concettualizzazione della coesione sociale ed accumulare conoscenza sul tema. Tale esigenza nasce dalla constatazione che la letteratura sulla coesione sociale ha carattere prevalentemente disomogeneo e poco strutturato con studi provenienti da diverse discipline e diversi ambiti (accademico, istituzionale, politico) che hanno impedito uno sviluppo di una teoria organica e coerente della coesione sociale.

Nel secondo capitolo, allo scopo di rendere maggiormente intelligibile la sezione empirica, vengono illustrate una serie di tecniche di analisi riferibili principalmente all'analisi multilivello e ai modelli ad equazioni strutturali multilivello. Oltre alla formalizzazione matematica, si mettono in luce i pregi, i limiti, la logica e i principi sottostanti a tali tecniche e si giustifica metodologicamente le situazioni di analisi in cui un ricorso ad un approccio multilivello è necessario. Tale capitolo risulta fondamentale all'interno dell'economia dell'intero lavoro in quanto le tecniche illustrate si configurano come altamente innovative all'interno del panorama internazionale soggette ancora a costante sviluppo, ricerca e dibattito metodologico. In particolare, la sezione dedicata ai modelli ad equazioni strutturali multilivello costituisce una delle prime pubblicazioni in lingua italiana, se non l'unica, presente sul tema.

Nel terzo e quarto capitolo del lavoro viene presentata l'indagine empirica. Innanzitutto, sulla base dei principali contributi teorici presenti in materia e dello schema concettuale sviluppato nella prima sezione, viene presentata la proposta di definizione operativa, teoricamente argomentata, del concetto di coesione sociale. Tale concetto multidimensionale viene concepito come composto da sette sotto-dimensioni costituenti.

Gli obiettivi dell'indagine empirica sono molteplici. L'analisi si prefigge di:

- testare empiricamente la validità della definizione operativa proposta;
- fornire una definizione operativa multilivello della coesione sociale con l'obiettivo di verificare se il costrutto svolge le stesse funzioni a livello individuale e aggregato, testandone pertanto l'isomorfismo a livello individuale e a livello aggregato;
- testare l'ipotesi che il modello di coesione sociale proposto sia invariante all'interno dei 29 Paesi presi in esame;
- specificare un modello ad equazioni strutturali multilivello nel tentativo di individuare i principali macro-fattori esterni che possono esercitare un'influenza sulla coesione sociale soffermandosi, in particolare, sugli aspetti legati alla diversità di etnie e alla sfera afferente l'ambito delle disuguaglianze all'interno di uno Stato.

La base di dati utilizzata riguarda l'indagine denominata *European Social Survey* che ha visto la partecipazione di 29 Paesi e 54672 individui.

Le tecniche di indagine utilizzate al fine di rispondere agli obiettivi cognitivi prefissati riguardano l'analisi fattoriale confermativa multilivello, l'analisi fattoriale confermativa multilivello di secondo ordine, i modelli ad equazioni strutturali multilivello e, infine, i modelli di mediazione ad equazioni strutturali multilivello.

Il modello di coesione sociale multilivello – composto da sette costrutti latenti riferiti alle sette sotto-dimensioni costituenti la coesione sociale individuate in fase di definizione operativa – gode di numerose proprietà che lo rendono interessante ed innovativo. Innanzitutto, viene mostrata l'equivalenza della struttura fattoriale del modello ai due livelli di analisi (individui, Paesi) illustrando in tal modo l'isomorfismo e l'equivalenza funzionale dei costrutti di primo e secondo livello. Inoltre, si evidenzia come tale modello sia invariante all'interno dei 29 Paesi considerati, ovvero il modello di coesione sociale individuato è esattamente lo stesso in ognuno dei 29 Stati conservandone la sua validità. Ciò evidenzia come i meccanismi che partecipano alla coesione sociale siano gli stessi a prescindere dai diversi contesti considerati.

Attraverso un ulteriore modello di analisi fattoriale multilivello di secondo ordine vengono individuati due fattori generali situati ai due livelli di analisi considerati, denominati *coesione sociale* (livello aggregato) e *inclusione sociale* (livello individuale). I due fattori sono responsabili della variazione dei sette costrutti di primo ordine. In particolare, si nota come un terzo della variabilità del fenomeno in esame sia da attribuirsi al livello aggregato – ovvero il Paese in cui si risiede contribuisce per il 35% della varianza del costrutto *coesione sociale*. Si evidenzia, quindi, come il contesto macro eserciti un'influenza più che rilevante sugli atteggiamenti e sui comportamenti individuali.

Infine, viene stimato un modello ad equazioni strutturali multilivello volto ad analizzare le relazioni sussistenti fra i sette costrutti latenti della coesione sociale e le sfere afferenti alla frammentazione etnica presente in uno Stato e alle disuguaglianze di reddito. Attraverso un modello di mediazione multilivello, si rileva come la *frammentazione etnica* non sempre eserciti un effetto negativo sulla coesione sociale. Piuttosto, tale relazione è spesso mediata da elementi afferenti alla sfera economica declinata nell'indagine in termini di disuguaglianze di reddito fra i cittadini di un Paese.

Si conclude sostenendo che, all'interno di una società, ampi livelli di diversità non necessariamente impattano negativamente sulla coesione sociale mettendone a rischio l'ordine sociale, piuttosto è il grado di disuguaglianza che sussiste fra soggetti che può produrre fenomeni di erosione della coesione sociale.

Il lavoro qui presentato costituisce un *unicum* nel panorama internazionale in quanto rappresenta l'unica indagine in cui si tenta di fornire una definizione multilivello della coesione sociale. In particolare, viene stabilito un parallelo fra i processi di tipo individuale e i processi aggregati attraverso l'individuazione di due costrutti isomorfi di ordine maggiore (*coesione sociale* ed *inclusione sociale*), responsabili dei meccanismi di produzione della coesione: i Paesi con maggior coesione sono il frutto di condizioni aggregate che producono negli individui atteggiamenti e comportamenti favorevoli all'incremento dell'inclusione aumentandone le interazioni positive e, a loro volta, tali interazioni interpersonali operano al fine di mantenere e riprodurre tali condizioni positive.

1. Il concetto di coesione sociale

1.1 La coesione sociale: un'introduzione teorica

Negli ultimi 20 anni, il concetto di coesione sociale è stato oggetto di crescente attenzione da parte di differenti settori come il mondo accademico e quello politico-istituzionale. Le molteplici motivazioni responsabili di tale interesse sono sostanzialmente riassumibili nelle domande fondamentali alle quali tale concetto permetterebbe di fornire una risposta. Come è possibile l'esistenza della società e il suo continuo riprodursi? Qual è l'elemento che permette ai singoli individui di costituirsi in società rendendo quest'ultima qualcosa di possibile?

Naturalmente, vi è un ulteriore ordine di motivazioni, anche di carattere strumentale, che ha determinato un'ampia diffusione di tale concetto soprattutto all'interno del mondo politico-istituzionale, dovuta all'estrema elasticità di tale concetto che lo rende adattabile ad ogni esigenza di *policy*.

Le domande sopra esposte ci riconducono direttamente agli albori delle scienze sociali. A ben vedere, costituiscono le stesse domande che a loro tempo si ponevano i padri della sociologia.

Durkheim è stato uno dei primi scienziati sociali ad occuparsi direttamente della coesione sociale. Nel periodo storico in cui egli visse la società fu attraversata da profondi cambiamenti strutturali; esattamente come sta accadendo attualmente. Ogni volta che l'assetto societario esistente viene minacciato da complessi mutamenti epocali emerge un problema di ordine sociale (coesione).

L'autore francese era essenzialmente interessato a capire come, all'interno di una società sconvolta da repentini cambiamenti – si pensi al passaggio da una società rurale ad una di tipo industriale – l'ordine sociale, nonché i legami, potessero essere mantenuti. Per Durkheim, gli stravolgimenti della società non necessariamente generano disordine sociale ed "anomia". Il sociologo francese considerava la società come un'unità composta da più parti fra loro interconnesse ed interdipendenti. Ogni parte, però, è un'entità a se stante, la cui sopravvivenza dipende dal funzionamento generale dell'intera unità. Nelle società preindustriali questo ordine è stato mantenuto grazie ad una "coscienza collettiva", ossia un insieme di credenze e sentimenti comuni, che creavano un sentimento d'appartenenza. Nelle società moderne, questa omogeneità non è più pensabile a causa della crescente complessità del sistema che genera una progressiva differenziazione delle funzioni. Nella teoria durkheimiana (1893) la coesione sociale è strettamente interconnessa ai concetti di solidarietà ed integrazione. In riferimento alla solidarietà, sono individuabili due diverse forme di tale concetto: una solidarietà meccanica ed una solidarietà organica. Entrambe sono collegate rispettivamente a due diversi tipi di società: la prima di carattere tradizionale, la seconda a carattere moderno.

La solidarietà meccanica si fonda sulla somiglianza fra individui che, al di là di pochi elementi di differenziazione, condividono lo stesso sistema valoriale, gli stessi orizzonti interpretativi, le stesse rappresentazioni.

La solidarietà organica, per alcuni versi, è l'esatto contrario di quella precedente in quanto si basa sulle differenze che sussistono fra gli individui. Tale forma di solidarietà sarebbe il frutto della divisione del lavoro: attraverso la differenziazione delle funzioni che ogni soggetto svolge si genera un processo che progressivamente produce sempre maggiori differenziazione fra gli individui. In un tale sistema – dove ognuno ha una sua specifica funzione – gli individui, pur essendo fra loro diversi, sono maggiormente interconnessi che in una società premoderna. In questo caso i soggetti sono funzionalmente connessi: ogni singola persona svolge una differente funzione rendendo gli individui

interdipendenti in quanto ogni entità ha bisogno di un'altra per funzionare. È tale interdipendenza che genera la coesione sociale.

Se la prima ondata di cambiamenti sociali produsse una visione ottimistica in Durkheim che vedeva in questi non una fonte di minaccia ma un meccanismo di produzione di diversificazione funzionale, la seconda ondata di cambiamenti sociali (1930) ha generato pessimismo. Parsons era consapevole dei rischi legati ad un eccessivo liberalismo che, secondo la sua visione, avrebbero generato le tragedie europee avvenute prima del 1945.

Parsons considera la società come un sistema complesso composto da altri sotto-sistemi interdipendenti mantenuti insieme da un processo di condivisione dei valori riprodotti attraverso la socializzazione (1949). Tale processo di condivisione dei valori costituisce il meccanismo che produce l'ordine sociale. Notiamo come si ponga in antitesi rispetto alla prospettiva durkheimiana dove il meccanismo riferito alla condivisione dei valori si caratterizza come l'attributo specifico delle società arcaiche.

Vale la pena ricordare come vi siano altre scuole di pensiero in riferimento all'ordine sociale. Il liberalismo di stampo classico (per una breve rassegna si veda Jenson 1998). Concepisce la società come una semplice somma di singoli individui. Il valore supremo è costituito dalla libertà dell'individuo. L'idea principale è che il buon funzionamento della società sia un sottoprodotto del comportamento degli individui. In base a tale prospettiva le istituzioni giocano un ruolo secondario nel favorire l'ordine sociale che è totalmente demandato al mercato, al comportamento degli individui e alle istituzioni private, come la famiglia, le associazioni, le istituzioni ecclesiastiche. L'intervento statale è visto come qualcosa che limita la libertà di scelta dell'individuo.

Nella prospettiva di de Tocqueville il ruolo svolto dalle associazioni costituisce il pilastro centrale su cui si basa la sua intera teorizzazione. L'azione volontaria produce due effetti complementari: un effetto interno che favorisce ed incoraggia un comportamento cooperativo fra i membri; un effetto esterno che si riversa sull'intero sistema di governo favorendo la cooperazione sociale.

In opposizione al liberalismo classico, le teorie della democrazia (Jenson 1998) considerano le istituzioni democratiche come fondamentali nell'assicurare il buon funzionamento di una società. Secondo questa prospettiva, l'ordine sociale sarebbe il frutto di due aspetti fondamentali: il primo riguarda il ruolo attivo giocato dalle istituzioni nel favorire una distribuzione equa delle risorse attraverso politiche sociali e di sostegno (assegni di disoccupazione, pensioni, assegni familiari, etc.); il secondo è relativo all'assicurare l'uguaglianza e l'equità attraverso politiche che favoriscano tale condizione (si pensi alla scuola pubblica) e che garantiscano il pieno accesso a tutti i cittadini al mercato del lavoro assicurando in tal modo a tutti l'accesso al reddito.

Questo breve excursus introduttivo evidenzia come l'interrogarsi sull'ordine sociale – sui meccanismi che rendono possibile la riproduzione della società, sui rapporti che sussistono fra individuo e società – sia da sempre una delle maggiori problematiche su cui gli scienziati sociali si pongono interrogativi sin dall'inizio delle scienze sociali.

1.2 Diverse definizioni del concetto di coesione sociale

Anche se lo sviluppo del concetto di coesione sociale risale al diciannovesimo secolo, non è attualmente rilevabile una definizione condivisa di tale concetto (Chan et al. 2006; Dicks et al. 2010). Come numerosi studiosi rilevano (Berger-Schmitt e Noll 2000; Berger-Schmitt 2002; Noll 2002; Jenson 1998; Jeannotte 2003; Bernard 1999; Beauvais e Jenson 2002), la coesione sociale pone numerose

sfide sia da un punto di vista concettuale che su un piano metodologico. A tale complessità si associa un generale stato di confusione su cosa si debba intendere per coesione sociale. Tale situazione è ben sintetizzata da Jenson che afferma: "Social cohesion is an ambiguous concept because it can be used by those seeking to accomplish a variety of things. It is sometimes deployed in rightwing and populist politics by those who long for the good old days when life seemed easier, safer, and less threatening. But social cohesion can also be used by those who fear the consequences of excessively marketised visions of the future. There is no question that those within Canada and much of the international policy community who evoke social cohesion do so because they fear the results of structural adjustments that ignore social and political needs. They are decidedly facing the future, not the past" (1998, 37).

Seguendo Chan et al. (2006), è possibile individuare due approcci principali alla coesione sociale. Il primo è un approccio di natura accademica con la maggioranza degli studi effettuati provenienti dalle discipline inerenti la sociologia e la psicologia sociale. All'inverso, il secondo tipo di approccio proviene dal mondo istituzionale sviluppato direttamente dai diversi *policymakers*.

Entrambe le prospettive scontano una serie di limitazioni e mancanze che progressivamente saranno messe in luce nei paragrafi che seguono. Per adesso ci basta dire che l'approccio accademico ha il forte limite di non essere stato in grado di sviluppare una visione comune e condivisa su cosa debba intendersi col termine coesione sociale. Mentre, l'approccio istituzionale ha l'ancor più grave limite nell'aver identificato la coesione sociale in base alle esigenze di *policy* delle singole istituzioni che di volta in volta si sono occupate di tale aspetto. Ciò ha prodotto uno stiramento semantico del concetto arrivando progressivamente ad indentificare la coesione sociale con gli urgenti problemi di natura contingente che i governi e le istituzioni si sono trovati via via a fronteggiare.

In base alla nostra prospettiva, sarebbero identificabili, in realtà, tre approcci diversificati alla coesione sociale. Il primo è diretta emanazione delle pubblicazioni ufficiali che periodicamente vengono rilasciate da diverse istituzioni (l'Unione Europea, l'Ocse, i diversi governi nazionali). Il secondo è stato sviluppato, invece, sia da analisti che da accademici partecipanti a diversi progetti di ricerca avente per oggetto la coesione sociale e direttamente finanziati dalle istituzioni a cui è stato fatto riferimento sopra. Infine, il terzo approccio è identificabile come un approccio puramente accademico.

Nei paragrafi che seguono, si proverà ad illustrare il dibattito scientifico relativo al concetto di coesione sociale cercando al contempo di metterne in luce gli aspetti positivi e quelli negativi. In seguito, basandoci sull'intero quadro concettuale e teorico pian piano delineato verrà proposta una nuova concettualizzazione della coesione sociale. In particolare, sarà proposto uno schema concettuale che ha l'obiettivo di porsi come strumento utile agli analisti, alle istituzioni e all'accademia capace di facilitare lo studio della coesione sociale sia da un punto di vista teorico che empirico in riferimento a tutte le fasi necessarie relative alla definizione, rilevazione e misurazione del concetto.

1.2.1 L'approccio istituzionale

L'approccio istituzionale, aspetto che ne costituisce anche il suo principale limite, è essenzialmente "guidato dai problemi" (*problem-driven approach*). La coesione diventa fondamentalmente una panacea per tutte le nuove forme di problemi che l'attuale società moderna si trova a fronteggiare. Tale prospettiva strumentale genera due ordini di problemi fra loro interconnessi. Innanzitutto, molti studi provenienti dall'ambito istituzionale si riducono ad essere semplicemente una lista molto corposa di indicatori volti alla rilevazione della coesione sociale senza nessun tipo di analisi preliminare volta

all'individuazione dei limiti concettuali e semantici della coesione e delle sue sotto-dimensioni costituenti. Inoltre, dal momento che la prospettiva assunta è di tipo strumentale, ossia volta a risolvere una serie di problematiche concrete, ogni istituzione ha sviluppato una propria concettualizzazione della coesione sociale in accordo alle proprie esigenze di *policy*. In altri termini, l'approccio istituzionale ha la tendenza ad indentificare e ad operativizzare la coesione sociale nei termini dei diversi problemi che le varie istituzioni e governi si trovano a fronteggiare e che assumono rilevanza all'interno del loro spazio di azione. In tal modo la prospettiva viene rovesciata. Nel caso in cui un governo si trovi a dover affrontare problemi relativi all'immigrazione o al multiculturalismo la coesione diventa un mezzo capace di risolvere tali problematiche attraverso aspetti riguardanti la condivisione di valori e la condivisione di un comune orizzonte di senso. Al contrario, se il problema consiste nella diffusa disuguaglianza che sussiste fra i cittadini, si esaltano aspetti diversi della coesione sottolineando l'importanza della riduzione delle disuguaglianze di reddito o, più in generale, del miglioramento delle condizioni di vita di un soggetto.

I maggiori contributi al concetto di coesione sociale provenienti dall'approccio istituzionale riguardano i documenti rilasciati dall'Unione Europea, dal Consiglio d'Europa e dall'Ocse (Jeannotte 2000).

Il concetto di coesione sociale dell'Unione Europea si basa sul modello europeo di società fondato su aspetti riguardanti la solidarietà, il sistema di protezione sociale e la continua regolazione e correzioni apportate ai "fallimenti del mercato" (Unione Europea 1996). Infatti, nel *report* dell'UE sulla coesione sociale del 1996 leggiamo che "a starting point would be to link social cohesion with the objectives of the European model of society which is founded on the notion of the social market economy [...] The solidarity dimension is given practical effect through universal systems of social protection, regulation to correct market failure and systems of social dialogue. The promotion of social cohesion requires the reduction of the disparities which arise from unequal access to employment opportunities and to the rewards in the form of income. Such inequality tends to have serious social consequences through the marginalisation of sections of society, such as the long-term unemployed, the young unemployed and the poor" (Unione Europea 1996, p.14).

Risulta essere chiaro, quindi, come la nozione di coesione sociale espressa dall'Unione Europea sia fortemente incentrata sugli aspetti legati più strettamente alla dimensione economica intesa soprattutto in termini di riduzione delle disuguaglianze di reddito. Nel *report* non viene fornita nessuna definizione diretta della coesione sociale; piuttosto, vengono identificati una serie di fattori che si pongono come elementi che possono minacciare i livelli di coesione sociale (ad esempio la disoccupazione, la povertà, l'esclusione sociale).

La concezione di coesione fornita dall'Ocse risulta essere più limitata, imperniata totalmente sugli aspetti collegati alla dimensione economica. Secondo l'Ocse ci sarebbe coesione laddove "la vita è caratterizzata da stabilità, continuità, prevedibilità e accesso sicuro al benessere materiale" (Jeannotte 2000). Ugualmente a quanto accade nel *report* dell'UE, i principali fattori di rischio sono indentificati nella disoccupazione, povertà e disparità di reddito.

Pur concordando con le due precedenti concettualizzazioni della coesione e affermando che tra gli obiettivi primari della coesione sociale vi siano la riduzione della povertà, della disoccupazione e delle disuguaglianze, il Consiglio d'Europa sostiene che l'obiettivo primario debba considerarsi il processo di ricostruzione del legame sociale fra individui e comunità e del legame politico fra cittadini e governi (Social, Health and Family Affairs Committee 1998).

Nessuna delle istituzioni citate fino ad ora fornisce una definizione esplicita su cosa debba intendersi con il termine coesione sociale. La loro strategia di analisi è totalmente rovesciata, concentrano, piuttosto, l'attenzione su una serie di fattori che possono erodere, o, viceversa, promuovere la

coesione sociale. L'analisi della lista dei principali fattori che possono compromettere la coesione sociale suggeriti dalle tre diverse organizzazioni in esame (tab.1.1) mostra un sostanziale disaccordo su cosa debba considerarsi effettivamente un fattore di rischio. Fra tutti, soltanto i fattori riferibili alla disoccupazione, alla disuguaglianza di reddito, all'esclusione sociale e all'esclusione dalla società dell'informazione sembrano mettere d'accordo i tre soggetti istituzionali.

Tab.1.1 – Fattori di rischio per la coesione sociale

<i>Threat</i>	<i>EU</i>	<i>OECD</i>	<i>CE</i>
<i>Political</i>			
Lack of a sense of European citizenship	√		
Political disenchantment		√	
The "rights deficit"	√		√
<i>Economic</i>			
Unemployment	√	√	√
Poverty and income inequality	√	√	√
Regional disparities	√		
<i>Social</i>			
Deterioration of the environment and quality of life	√		
Social exclusion	√	√	√
Exclusion from the Information Society	√	√	√
<i>Cultural</i>			
Cultural diversity and demographic change		√	
Changing values regarding work and society		√	
Influence of American culture			√
Shift from "culture of collective security" to "culture of individual opportunity"			√

Source: Jeannette 2000.

L'Ocse, in aggiunta ai fattori sopra illustrati, considera le diversità culturali e i cambiamenti demografici come fattori di rischio capaci di minare la coesione di un popolo. Infatti, pur riconoscendo al fenomeno dell'immigrazione la capacità di rigenerare una società, l'Ocse considera la crescente disparità fra cittadini e immigrati una minaccia seria per la stabilità sociale.

Per l'Unione Europea, le disuguaglianze sociali ed economiche rappresentano il pericolo maggiore per la coesione sociale di una Nazione. L'UE identifica nell'esclusione sociale – intesa non solo in termini riconducibili esclusivamente alla povertà economica ma come impossibilità ad esercitare pienamente i diritti sociali, culturali e politici – uno dei maggiori elementi destabilizzanti per una società. Inoltre, tra i problemi più rilevanti, l'UE inserisce la riduzione delle disparità fra gli Stati Membri dell'UE – vale la pena notare come tale aspetto costituisca al contempo uno dei principali obiettivi dell'intera politica dell'UE (Unione Europea 1996; Geddes 1995).

Il consiglio d'Europa condivide le preoccupazioni dell'Unione Europea. Nuovamente, l'esclusione sociale, intesa come espulsione da almeno uno dei tre poli dell'integrazione – Stato, mercato economico e società civile – rappresenta uno dei principali fattori che possono mettere a rischio la stabilità sociale. Interessante notare come il Consiglio d'Europa consideri esplicitamente come una minaccia per la coesione sociale in Europa l'aspetto collegato ad una sorta di importazione nel vecchio continente della cultura americana con conseguente passaggio ad una cultura di stampo individualista che possa minare la basa valoriale esistente (Social, Health and Family Affairs Committee 1998; The European Task Force on Cultural and Development 1997).

Le tre organizzazioni, oltre a stilare i principali fattori che possono logorare quote importanti di coesione sociale, forniscono anche i principali elementi che potrebbero, all'inverso, favorire lo sviluppo

della coesione sociale (tab.1.2). Si può notare come la lista dei fattori di promozione presentata in tab.1.2 sia significativamente meno nutrita di quella presentata in tab.1.1. Non a caso, la letteratura sulla coesione sociale, soprattutto nella sua fase iniziale, tendeva a concentrarsi esclusivamente sui fattori che potessero esercitare un effetto negativo sulla coesione piuttosto che ricercare gli elementi che potessero in qualche modo promuoverla (Jenson 1998; Beauvais e Jenson 2002).

L'Ocse individua il maggior numero di fattori di promozione. Tra questi, il più importante viene considerato il fattore relativo alla capacità di assicurare la presenza di forti mercati economici competitivi associati alla presenza di un forte tessuto sociale bilanciando il dinamismo e la sicurezza sociale al fine di evitare la stagnazione e la frammentazione sociale. Quest'ultimo, più che un fattore di promozione ci sembra un complesso obiettivo di *policy* esplicitamente dichiarato. È significativo far notare come, dei tre Istituti presi in considerazione, soltanto l'Ocse consideri l'accettazione e il rispetto delle diversità come un fattore chiave che possa contribuire positivamente ad elevare le quote di coesione sociale (Michalski, Miller e Stevens 1997).

L'Unione Europea identifica soltanto due principali fattori che agirebbero positivamente sulla coesione sociale: il "modello europeo di sviluppo e solidarietà" che si basa sulla lotta alla povertà, alle disuguaglianze attraverso meccanismi perequativi e l'intervento diretto dello Stato; un senso d'appartenenza ad una comune identità europea che, oltre a favorire un sentimento di cittadinanza condivisa, contribuisca alla riduzione della disoccupazione attraverso la promozione di maggiori scambi fra Stati europei (Jeannotte 2000).

Allo stesso modo, il Consiglio d'Europa mette al centro la solidarietà basata sul modello europeo di sviluppo ma, rispetto all'UE, individua numerosi altri fattori. Tra questi vi è certamente il rafforzamento delle tecnologie dell'informazione che permettono di intrattenere legami più intensi fra i vari Stati Membri. Un altro fattore di sicuro rilievo si riferisce al "rinnovo della società civile" inteso in termini di rafforzamento dell'aspetto legato all'associazionismo (Consiglio d'Europa 1997; Jeannotte 2000).

Tab.1.2 – Fattori di promozione della coesione sociale

<i>Factor</i>	<i>EU</i>	<i>OECD</i>	<i>CE</i>
<i>Political</i>			
Solidarity - the "European model"	√		√
<i>Economic</i>			
Capacity to balance competitive markets with strong social fabric		√	
Dematerialization and the "culturization" of the economy		√	
<i>Social</i>			
Capacity for dialogue and cooperation		√	√
Societal flexibility and demographic change		√	
Renewal of civil society			√
<i>Cultural</i>			
Culture and a sense of belonging	√		
Openness to diversity		√	
<i>Technological</i>			
Potential of information technologies			√

Source: Jeannotte 2000.

Riassumendo, l'Ocse intende la coesione sociale come un elemento direttamente connesso al benessere: la coesione sociale si pone in tal modo come un antidoto ai repentini mutamenti del mercato economico e alla crescente flessibilità del mercato del lavoro. Mentre, sia per l'Unione

Europea che per il Consiglio d'Europa la coesione sociale si riferisce essenzialmente agli aspetti legati alla riduzione delle disuguaglianze che sussistono fra i cittadini nonché alla lotta volta alla riduzione dell'esclusione sociale e alla diminuzione dei tassi di disoccupazione. Possiamo notare, quindi, come la concezione dei tre Istituti della coesione sociale sia fondamentalmente basata esclusivamente sulla dimensione economica del concetto in esame.

A conferma, l'*European Committee for Social Cohesion* ha ulteriormente rafforzato tale idea sostenendo come la coesione sociale si possa definire in termini di capacità da parte di una società di assicurare il benessere di tutti i cittadini riducendone al minimo le disparità ed evitando fenomeni di polarizzazione sociale (2004). Per l'*European Committee for Social Cohesion* una società coesa è contraddistinta dalla presenza di una comunità di individui liberi in cui vige il mutuo supporto ed in cui tutti partecipano al raggiungimento di obiettivi comuni attraverso il ricorso a mezzi democratici (2004). Nuovamente, il *focus* dell'attenzione è volto agli aspetti collegati alla riduzione delle disuguaglianze e allo scongiurare fenomeni di polarizzazione.

Nel sesto Rapporto sulla Coesione Economica, Sociale e Territoriale (2014) la Commissione Europea afferma che gli obiettivi riguardanti le politiche di coesione sono da intendersi come principalmente volte alla riduzione delle disuguaglianze sociali, economiche e territoriali fra gli Stati Membri fornendo il supporto necessario al fine di colmare il divario fra le diverse regioni europee. Essenzialmente, nel sesto Rapporto sulla Coesione Economica, Sociale e Territoriale l'approccio generale alla tematica relativa alla coesione sociale non è mutato, con una prospettiva totalmente appiattita su una dimensione esclusivamente economica in cui la coesione viene considerata un fattore capace di ridurre il fenomeno della disoccupazione, l'esclusione sociale e le disparità fra gli Stati Membri ma senza mai definire cosa debba intendersi con l'espressione "coesione sociale".

Il Government of Canada's Policy Research Sub-Committee si pone nei confronti della tematica generale ricorrendo ad una diversa prospettiva. Innanzitutto, diversamente da quanto visto in precedenza, ne fornisce una delle poche definizioni esplicite: "the ongoing process of developing a community of shared values, shared challenges and equal opportunity within Canada, based on a sense of trust, hope and reciprocity among all Canadians" (Berger 1998). Notiamo, innanzitutto, come la definizione implichi una concezione dinamica della coesione sociale, definita appunto come processo continuo (*ongoing process*); inoltre, poggia su basi concettuali totalmente differenti rispetto a quelle viste in precedenza. Se precedentemente l'aspetto principale riguardante la coesione veniva individuato nella riduzione delle disuguaglianze economiche e sociali, il Governo Canadese, coerentemente con gli aspetti problematici interni alla società canadese, pone l'attenzione nel sottolineare gli aspetti culturali (condivisione di valori, fiducia, reciprocità) legati alla coesione sociale. In tale aspetto risulta evidente una delle maggiori limitazioni dell'approccio istituzionale: la tendenza a definire la coesione sociale nei termini delle problematiche sociali che le differenti istituzioni devono risolvere.

Il Governo Francese mostra lo stesso orientamento del Governo canadese ritenendo la coesione sociale assimilabile ad un processo piuttosto che ad una condizione (concezione dinamica vs. concezione statica). Pertanto, il *Commissariat général du Plan* definisce la coesione come un insieme di processi sociali che favoriscono l'incremento di un senso di appartenenza alla stessa comunità e la consapevolezza da parte degli individui di essere riconosciuti come membri di tale comunità (Jenson 1998). Nuovamente, gli elementi centrali della definizione sono individuabili negli aspetti culturali del concetto come il sentimento di appartenenza, la fiducia, il riconoscimento.

Il Governo Inglese fornisce, all'opposto, una concettualizzazione della coesione sociale che risulta essere più in linea con quelle fornite dall'UE e dal Consiglio d'Europa. Una comunità viene definita

coesa quando è possibile identificarne all'interno i seguenti elementi: gli individui con un differente status socio-economico godono delle stesse opportunità; sussistono forti legami sociali fra soggetti con differente status socio-economico; vi è un diffuso senso di appartenenza alla comunità; infine, le differenze culturali non sono giudicate negativamente ma considerate piuttosto un valore aggiunto (House of Commons 2004).

In definitiva, l'approccio istituzionale ha la caratteristica di essere un *problem-driven approach*. Come si sostiene in Chan et al. (2006), il discorso intorno alla coesione sociale si pone essenzialmente come reazione alle innumerevoli nuove problematiche che si pongono come una sfida alla nostra società. Questo orientamento guidato dalle esigenze di *policy* è chiaramente visibile quando analizziamo contemporaneamente le definizioni della coesione sociale fornite dagli Istituti visti sopra e le problematiche che le stesse Istituzioni si trovano a fronteggiare. Per esempio, in Canada gli aspetti legati al multiculturalismo costituivano una delle principali tematiche dell'agenda di governo così come la coesistenza fra parlanti di lingua inglese e quelli di lingua francese. Così, in Canada, la coesione sociale diventa essenzialmente una questione relativa a tematiche inerenti la condivisione di valori, l'appartenenza ad una stessa comunità, mutuo supporto e fiducia. In Europa, dopo l'entrata nella Comunità Europea di diversi Stati caratterizzati da condizioni economiche arretrate, la coesione sociale è stata intesa come riduzione delle disparità e omogeneizzazione delle condizioni di vita di tutti i cittadini europei. Numerosi aspetti della politica europea nonché ingenti risorse sono totalmente indirizzate al raggiungimento di tale obiettivo. In tal modo la prospettiva d'analisi viene rovesciata e la coesione sociale viene spesso assimilata alle problematiche principali presenti in una società.

Strettamente interconnesso a tale aspetto, si aggiunge un'ulteriore debolezza all'interno dell'impianto concettuale dell'approccio istituzionale. Molti fattori sono considerati contemporaneamente sia come un obiettivo da raggiungere sia come mezzo per raggiungere tale obiettivo. Il modello europeo di sviluppo e solidarietà (tab.1.2) costituisce contemporaneamente sia un fattore di promozione della coesione sociale sia uno degli obiettivi cardine delle politiche europee di sviluppo. Allo stesso modo la dimensione relativa alla condivisione dei valori o alla promozione di uguali condizioni di vita fra gli individui rappresentano simultaneamente sia fattori capaci di promuovere la coesione sociale sia obiettivi da raggiungere, grazie alla coesione sociale.

Quello istituzionale potrebbe pertanto essere definito "approccio mezzi-fini" dove i mezzi e gli obiettivi da raggiungere attraverso gli strumenti individuati sono spesso intrecciati a tal punto da confondere gli antecedenti e i conseguenti: in tal modo i "percorsi causali" fra i diversi fattori risultano difficilmente individuabili. Come meglio vedremo, ciò produce un altro fenomeno: il continuo confondere e mischiare i fattori che esercitano un'influenza sulla coesione e le sue dimensioni costituenti (Chan et al. 2006).

1.2.2 L'approccio policy-oriented

L'approccio orientato alle politiche proviene essenzialmente da ricercatori o accademici collegati a svariati istituti di ricerca. Si è preferito considerare l'approccio politico e quello istituzionale come due orientamenti distinti in quanto riteniamo che il primo sia scientificamente e concettualmente più accurato con alcuni contributi che rappresentano tentativi di concettualizzazione della coesione fra i più efficaci e autorevoli. Sebbene molti dei contributi che saranno presentati in questo paragrafo siano stati sviluppati da soggetti appartenenti alle istituzioni accademiche, si è scelto di includerli

nell'approccio orientato alla politica in quanto molti di loro hanno partecipato a progetti di ricerca direttamente finanziati da alcune Istituzioni con obiettivi di *policy* espressamente dichiarati.

Vi sono stati numerosi tentativi di concettualizzare la coesione sociale, certamente uno dei più autorevoli proviene da un'indagine realizzata da Jenson, professore all'Università di Montreal, per il Canadian Policy Research Networks (Jenson 1998). L'autrice considera la coesione sociale come concetto multidimensionale composto da cinque sotto-dimensioni:

1. Belonging – Isolation: riferita alla condivisione di valori e alla presenza di un'identità collettiva.
2. Inclusion – Exclusion: relativa all'aspetto collegato alle pari opportunità di accesso al mercato economico da parte di tutti i cittadini.
3. Participation – Non-involvement: riguardante la partecipazione intesa sia nella sua dimensione politica che sociale.
4. Recognition – Rejection: concernente gli elementi riguardanti il rispetto e la tolleranza delle diversità.
5. Legitimacy – Illegitimacy: riferita alla mantenimento della legittimità delle istituzioni che agiscono come soggetto mediatore riducendo il conflitto sociale.

Le cinque dimensioni rappresentano dei *continua*, la coesione sociale di una società è data dalla sua posizione su ogni *continuum*.

Coerentemente con le necessità della politica canadese e le condizioni generali della società canadese, Jenson fornisce una concettualizzazione della coesione sociale concentrata prevalentemente sulla dimensione culturale. Nella nozione di coesione fornita da Jenson, gli aspetti collegati alla condivisione di valori, ad un sentimento di appartenenza ad una comunità collettiva, al rispetto delle differenze in una società multiculturale si impongono come fattori fondamentali in una società che possa definirsi come coesa. Come Chan et al. fanno notare (2006), almeno due delle sotto-dimensioni della coesione sociale individuate da Jenson – nello specifico l'inclusione e il riconoscimento – si caratterizzerebbero per non essere dimensioni costituenti del concetto. Infatti, "dimensioni costituenti" e "condizioni" sono due nozioni concettualmente distinte. L'inclusione che, nella formulazione fornita da Jenson, si riferisce alle pari opportunità all'interno del mercato economico non si caratterizza come una dimensione che va a costituire il concetto di coesione sociale; piuttosto, si pone come fattore (condizione) che può esercitare un'influenza sulla coesione. Non sarebbe quindi una sotto-dimensione del concetto ma un fattore esterno che può agire più o meno positivamente sulla coesione di uno Stato. Piuttosto che considerare l'inclusione economica una dimensione costituente della coesione sociale, come suggeriscono Chan et al. (2006), si dovrebbe piuttosto verificare empiricamente le relazioni che sussistono fra la coesione sociale e la parità di opportunità fra tutti i cittadini. Secondo Chan et al. (2006), non è nemmeno certo che l'uguaglianza di opportunità fra gli individui sia una condizione necessaria al fine della promozione di quote crescenti di coesione sociale: vi sono numerosi esempi nella storia dell'umanità di società fortemente coese nelle quali le pari opportunità erano una nozione totalmente sconosciuta.

Introducendo una distinzione fondamentale fra aspetti formali e sostanziali all'interno della concettualizzazione di Jenson, Bernard (1999) individua una mancanza all'interno della nozione di coesione appena illustrata. L'autore propone uno schema della coesione sociale integrando le dimensioni suggerite da Jenson attraverso l'aggiunta di un'ulteriore dimensione. Infatti, introducendo una distinzione fra diverse sfere di attività – economica, politica e culturale – aggiunge una sesta dimensione alle cinque viste sopra che definisce "uguaglianza-disparità" che si pone come aspetto

sostanziale della sfera di attività economica. Nella formulazione originariamente presentata da Jenson, come detto, tale aspetto era assente con la presenza del solo aspetto formale della sfera di attività economica rappresentato dalla dimensione "inclusione-esclusione".

Gli stessi problemi presenti nella formulazione di Jenson si ritrovano in un'altra proposta canadese. Duhaime et al. (2004), rifacendosi alla distinzione operata da Durkheim fra solidarietà organica e solidarietà meccanica, afferma che la coesione sociale è composta da due dimensioni: l'accesso alle istituzioni economiche e governative (che favorisce la solidarietà organica) e l'accesso alle relazioni faccia a faccia sia familiari che all'interno della comunità (che produce la solidarietà meccanica). Nella prima dimensione vi sono i legami sociali prodotti e riprodotti dalla partecipazione al mercato del lavoro, dalla partecipazione politica e dall'accesso ai servizi pubblici. La seconda dimensione, all'inverso, comprende i legami sociali generati dalla partecipazione ad attività familiari e comunitarie che non implicino uno scambio economico.

L'autore, al fine di coprire semanticamente le due dimensioni concettuali individuate, sviluppa sei set di indicatori:

1. Presenza di capitale sociale: relativo alla fiducia nelle istituzioni e alla partecipazione ad associazioni;
2. Stabilità demografica: riguardante la mobilità degli individui, il tasso di crescita della popolazione, le motivazioni soggettive relative al rimanere o uscire da una comunità;
3. Inclusione sociale: riferito all'accesso a reti informali che forniscono supporto affettivo, cognitivo e materiale;
4. Inclusione economica: relativo alla partecipazione al mercato del lavoro;
5. Qualità della vita della comunità: include la soddisfazione relativa ad una serie di servizi e ad un sentimento di sicurezza soggettivo all'interno della comunità;
6. Qualità della vita individuale: riferito alla salute e alla soddisfazione e felicità soggettiva.

Indubbiamente, il contributo fornito da Duhaime et al. risulta essere uno dei più accurati e rappresenta sicuramente un *framework* molto dettagliato al fine della misurazione del concetto di coesione sociale. Tuttavia, secondo Chan et al. (2006) vi sono numerosi punti deboli in questa formulazione. Innanzitutto, non risultano chiare le relazioni che sussistono fra alcuni set di indicatori (in particolare il set 2, 5 e 6) e le due macro dimensioni riferite alle due forme di solidarietà. La stabilità demografica e la qualità della vita individuale e comunitaria non costituiscono in nessun modo forme di solidarietà. Secondariamente, se la coesione sociale è intesa da Duhaime et al. come "access to formal economic and governmental institutions" e "access to family and community-based, face to face relations" (2004, 301) soltanto i set numero 1, 3 e 4 dovrebbero apparire nella formulazione fornita. Infine, ritroviamo la confusione fra ciò che può definirsi coesione e ciò che si pone come elemento capace di influenzarla. I set numero 2, 4, 5 e 6 non rappresentano dimensioni costituenti della coesione sociale ma sono elementi che possono contribuire a favorire lo sviluppo di coesione sociale (Chan et al. 2006). Nella proposta di concettualizzazione della coesione sociale fornita da Berger-Schmitt (2000) e Noll (2002), la dimensione riferita alle disuguaglianze – riduzione delle disparità fra individui – diventa il pilastro concettuale su cui poggia l'intera teorizzazione. Gli autori propongono di suddividere il concetto generale di coesione sociale in due sotto-dimensioni che allo stesso tempo costituiscono due obiettivi di *policy* da raggiungere. La prima sotto-dimensione riguarda la riduzione della disuguaglianza, la seconda si riferisce al capitale sociale. La dimensione relativa alla disuguaglianza riguarda "the goal of promoting equal opportunities and reducing disparities and divisions within a society. This also

includes the aspect of social exclusion" (Berger-Schmitt 2002, 406). La dimensione riguardante il capitale sociale è definita come "the goal of strengthening social relations, interactions and ties and embraces all aspects which are generally considered as the social capital of a society" (Berger-Schmitt 2002, 406).

Berger-Schmitt (2002) suggerisce di suddividere ciascuna sotto-dimensione in ulteriori tre componenti. La dimensione relativa alla disuguaglianza comprende le seguenti componenti:

1. Disparità regionali (regioni dell'UE)
2. Uguaglianza di opportunità
3. Esclusione sociale.

La dimensione riguardante il capitale sociale invece include le seguenti componenti:

1. Relazioni sociali ed attività svolte all'interno di gruppi primari ed associazioni
2. Qualità delle relazioni sociali
3. Qualità delle istituzioni.

La concettualizzazione della coesione sociale appena illustrata è stata operativizzata nel contesto del progetto che va sotto il nome di European System of Social Indicators (EUSI).

Le due principali dimensioni che compongono la coesione sociale, contestualmente alle sei componenti complessive, sono state operativizzate all'interno di ciascuno dei 14 obiettivi di *policy* individuati nell'ambito del progetto denominati "*life domains*".

Una delle principali limitazioni attenenti alla formulazione proposta da Berger-Schmitt (2000) e Noll (2002) riguarda il fatto che la prospettiva adottata si caratterizza per essere una prospettiva mezzi-fini (Chan et al. 2006). La coesione sociale viene definita nei termini delle condizioni o obiettivi – più capitale sociale e meno disuguaglianza – che si vogliono raggiungere e che possono contribuire ad ampliare le quote di coesione sociale. Nuovamente, vi è una confusione su ciò che costituisce coesione sociale e ciò che, all'inverso, è un fattore influenzante. La dimensione della disuguaglianza non costituisce una dimensione propria alla coesione sociale ma, più plausibilmente, si pone come un fattore capace di esercitare un'influenza sul concetto in esame. Tale relazione non può essere semplicemente assunta ma necessita di essere empiricamente testata.

Indubbiamente, l'approccio orientato alle politiche ha il forte merito di avere riportato al centro dell'attenzione il concetto di coesione sociale a lungo dimenticato. Inoltre, come si vedrà più avanti, i tentativi espliciti di operativizzare il concetto di coesione sociale provengono prevalentemente da questo filone di studi piuttosto che dall'approccio puramente accademico. Oltre a tali meriti, abbiamo visto come vi siano anche debolezze collegate a tale prospettiva di analisi.

Tutte le principali limitazioni nascono essenzialmente dal fatto che l'orientamento adottato è di tipo *problem-driven*. Vi è la tendenza ad operativizzare la coesione sociale nei termini degli obiettivi di *policy* che ci si prefissa di raggiungere (riduzione della disuguaglianza, costruzione di una comunità dove all'interno viga la condivisione di valori). Tale aspetto conduce inesorabilmente a mescolare e confondere ciò che è proprio della coesione sociale – le sue dimensioni costituenti – e le condizioni che si suppongono possano influire sulla coesione rispecchiando allo stesso tempo l'orientamento politico dell'intero discorso sulla coesione sociale (Dickes et al. 2010; Chan et al. 2006). L'orientamento politico è ravvisabile anche nella moltitudine di definizioni che sono state fornite nel corso degli anni: il modo in cui la coesione sociale viene definita dipende dalle problematiche maggiormente presenti

in quella data società (per esempio multiculturalismo in Canada, disparità in Europa). Pertanto, la coesione sociale assume l'aspetto di un costrutto che apparentemente può risolvere i problemi più urgenti presenti all'interno delle società moderne; una sorta di parola-chiave che può dirimere tutte le controversie della modernità (in realtà tale orientamento è più marcato nell'approccio istituzionale visto sopra). Secondo Chan et al. (2006), il discorso dei decisori politici sulla coesione sociale è diventato "a 'catchword' for incorporating the most pressing issues of the day: unemployment, poverty, discrimination, exclusion, disenchantment with politics, together with any problems that a policymaker sees fit" (2006, 288).

1.2.3 L'approccio accademico

I contributi alla coesione sociale relativi all'approccio accademico provengono principalmente dalle discipline inerenti la sociologia e la psicologia sociale. Nel 1896, Gustave Le Bon, psicologo sociale francese, fornisce una prima spiegazione del comportamento collettivo. Secondo Le Bon (1895), le folle godono di una propria esistenza ed esercitano un'influenza diretta sui singoli membri. Le folle possono ispirare emozioni intense e azioni irrazionali. La solidarietà che può nascere all'interno delle singole folle deriva sostanzialmente dalla sua "uniformità di azione" la quale è generata da una forma di comportamento caratterizzato dall'anonimato e dal contagio.

La teoria del contagio di Le Bon è uno dei primi precursori del concetto di coesione sociale. Nello stesso periodo, Durkheim (1897) indaga la relazione che sussiste fra la coesione sociale e le forme di suicidio. Nelle comunità più coese – società con legami sociali più forti – gli individui tendono a commettere atti suicidari in misura minore.

Anche se il concetto di coesione sociale è presente fin dalla nascita delle scienze sociali, gli studiosi, nel tempo, hanno preferito orientarsi verso altri concetti fortemente collegati alla coesione. Concetti come l'integrazione sociale, la solidarietà, il capitale sociale hanno goduto di maggior fortuna all'interno dell'ambito accademico. Ciò non vuole dire che non sussistano contributi interessanti in materia; piuttosto, significa che manca una teoria condivisa, organica e sistematica della coesione sociale.

Diversamente dai due approcci visti in precedenza che tendono a considerare la coesione sociale come un concetto dinamico, l'approccio accademico si caratterizza per la preferenza di definizioni statiche¹. Gli studi sulla coesione sociale possono essere anche suddivisi tenendo in considerazione il livello di analisi nel quale il concetto viene considerato. La coesione può essere concettualizzata come una proprietà sistemica e la raccolta dei dati avviene ad un livello aggregato. Viceversa, seguendo l'individualismo metodologico, sebbene la coesione venga ancora intesa come proprietà predicabile di un aggregato, viene considerata l'effetto degli atteggiamenti e dei comportamenti individuali in diverse sfere dell'agire umano.

Rajulton et al. (2007) ricorrono all'uso di misure aggregate per misurare la coesione sociale. Gli autori individuano tre sfere nelle quali la coesione può essere presa in considerazione – sociale, economica e politica. Negli ambiti così identificati fanno rientrare le dimensioni della coesione sociale individuate da Jenson con l'integrazione fornita da Bernard. Quindi, individuano sei dimensioni in tutto: inclusione ed eguaglianza nella sfera economica; riconoscimento ed appartenenza nel dominio sociale; legittimità

¹ Le definizioni dinamiche considerano la coesione sociale come un processo volto al raggiungimento di alcuni obiettivi; le definizioni statiche intendono la coesione sociale come una condizione generale (*state of affair*) di una determinata società.

e partecipazione nell'ambito politico. Anche nella prospettiva fornita da Rajulton et al. (2007) la sfera economica intesa come uguaglianza/disuguaglianza fra gli individui viene considerata come una dimensione costituente della coesione sociale e non come un fattore esogeno.

Uno dei principali contributi al dibattito sulla coesione sociale ci viene fornito da Lockwood (1999). Partendo dal concetto di integrazione sociale, Lockwood teorizza l'esistenza di due assi concettuali. Il primo riguarda l'ordine istituzionale considerato ad un livello macro della società ed è rappresentato dal *continuum* integrazione civica-corrruzione civica. Il secondo asse si concentra sulle reti di relazioni primarie ed è costituito dal *continuum* coesione sociale-disfacimento sociale. I due assi rappresentano i due livelli dell'integrazione riguardante le relazioni tra individui che possono essere di tipo conflittuale o armonioso (Lockwood 1992). Più precisamente, la coesione sociale è relativa alle relazioni primarie e secondarie all'interno di una comunità locale. Lockwood ha proposto un possibile set di indicatori utili a rilevare la coesione sociale. Tra questi troviamo la partecipazione politica, il supporto fornito alla democrazia, la presenza di forme di estremismo politico, diritti sociali in riferimento all'asse relativo all'integrazione civica e grado di associazionismo, criminalità, presenza di famiglie disfunzionali per quanto riguarda la dimensione relativa alla coesione sociale (Lockwood 1999).

Come Chan et al. fanno notare (2006), gli indicatori utilizzati da Lockwood lasciano intendere che l'approccio utilizzato dall'autore risulta essere esclusivamente negativo – focalizzato sulle cosiddette patologie sociali – e nessun riferimento viene fatto ad indicatori positivi. Inoltre, alcune misure proposte sembrano caratterizzarsi maggiormente come condizioni (fattori) che possono contribuire ai livelli di coesione sociale piuttosto che dimensioni costituenti della coesione sociale. Per esempio, i livelli di criminalità potrebbero essere una conseguenza dovuta ad un basso grado di coesione all'interno di una comunità.

Bollen e Hoyle (2001), considerando la coesione come una proprietà situata non al livello della società ma come una proprietà dei gruppi sociali, introducono la distinzione fra due differenti prospettive relative alla coesione: oggettiva e percepita.

La prima si riferisce alle caratteristiche di un gruppo considerato nel suo insieme ed è rilevata tramite indicatori compositi derivati dal giudizio fornito dai singoli membri relativamente alla vicinanza o lontananza in riferimento agli altri componenti. La seconda prospettiva riguarda, invece, la percezione dei singoli membri in riferimento al loro *status* generale all'interno del gruppo.

Whelan e Maître (2005) introducono una distinzione fondamentale. Gli autori suggeriscono che la coesione sociale può essere analizzata a tre diversi livelli di analisi: micro, meso e macro. Il livello micro si riferisce a dimensioni relative alla fiducia interpersonale, la forza dei legami all'interno della famiglia e dei legami primari in generale e il rischio di essere escluso da tali reti relazionali. Il livello meso riguarda la forza delle relazioni all'interno dei gruppi secondari, rapporti di vicinato, reti lavorative e relazioni con differenti gruppi etnici. Il livello macro concerne sentimenti di appartenenza alla comunità, il rispetto delle differenze e la legittimità ed efficacia delle istituzioni. Se da un punto di vista concettuale il lavoro di Whelan and Maître è ben strutturato, la stessa cosa non si può dire dell'aspetto empirico. La scelta degli indicatori volti alla misurazione della coesione sociale non è pienamente soddisfacente. Infatti, gli indicatori selezionati per rilevare il concetto all'interno dei tre livelli di analisi non coprono interamente queste tre dimensioni. Gli indicatori inclusi nella definizione operativa della coesione sociale si riferiscono alla percezione della qualità dei servizi pubblici, all'atteggiamento nei confronti del sistema di *welfare*, all'alienazione intesa come generale sfiducia nei confronti della società intera, alla percezione del conflitto fra differenti gruppi sociali e alla fiducia interpersonale. Con

l'eccezione della fiducia interpersonale e la percezione del conflitto fra diversi gruppi sociali², tutti gli altri indicatori sono relativi alla rilevazione del concetto all'interno della dimensione macro. Quindi, la definizione operativa proposta da Whelan and Maître sembra poco affidabile dal momento che due intere dimensioni, livello micro e meso, sono rilevate soltanto attraverso un indicatore.

Più accurata è la concettualizzazione della coesione sociale proposta da Chan et al. (2006). Gli autori sono tra i pochi a fornire un'esplicita definizione della coesione che viene considerata "a state of affairs concerning both the vertical and the horizontal interactions among members of a society, as characterized by a set of attitudes and norms that include trust, a sense of belonging, and the willingness to participate and help, as well as their behavioural manifestations" (Chan et al. 2006, 290). La definizione fornita rende esplicito che la coesione sociale è considerata una proprietà della società anche se la coesione può essere misurata ricorrendo a dati individuali. Considerare la coesione come una proprietà significa allo stesso tempo non intenderla come un processo: quindi non esistono, nemmeno teoricamente, livelli ideali o massimi di coesione sociale.

Chan et al. (2006) forniscono uno schema della coesione sociale direttamente derivato dalla definizione fornita sopra. Lo schema include una dimensione orizzontale, riferita alle relazioni all'interno della società che sussistono fra individui e gruppi, e una dimensione verticale, focalizzata sulle relazioni vigenti fra stato e suoi cittadini. Lo schema prende in considerazione anche due componenti. La componente soggettiva riguarda la sfera degli atteggiamenti mentre la componente oggettiva si riferisce all'ambito afferente i comportamenti.

Quindi Chan et al. (2006) considerano in tutto quattro dimensioni e per ogni dimensioni suggeriscono una serie di possibili indicatori, riportati in tab.1.3 suddivisi per dimensione e componente di appartenenza.

Tab.1.3 – Schema della coesione sociale proposto in Chan et al. (2006)

	Componente soggettiva	Componente oggettiva
Dimensione orizzontale	-general trust with fellow citizens -willingness to cooperate -sense of belonging	-social participation -voluntarism
Dimensione verticale	-trust in public figures -confidence in political institutions	-Political participation

Il maggior punto di forza della formulazione della coesione sociale appena illustrata si riferisce soprattutto al fatto che gli autori, consapevolmente, hanno escluso la dimensione economica considerandola un aspetto che non costituisce una dimensione fondante della coesione sociale, piuttosto un fattore esogeno di influenza.

Un'altra interessante proposta ci viene da Dickes et al. (2010) e Dickes e Valentova (2013) che propongono una definizione di coesione sociale basata sull'integrazione della teoria di Bernard (1999) e la proposta di Chan et al. (2006). Gli autori mostrano che lo schema proposto da Chan et al. (2006) si

² Non è chiaro se tale indicatore si riferisce alla dimensione macro o meso. Tenendo conto che il livello meso è focalizzato sulla forza delle relazioni all'interno dei gruppi secondari, sembra doversi riferire maggiormente al livello macro.

sovrappone sostanzialmente alla concettualizzazione di Bernard (1999, vedi sopra) con l'eccezione che, nel contributo di Chan et al., gli autori eliminano l'ambito riferito alla dimensione economica. Eliminando allo stesso modo la dimensione economica dal computo delle dimensioni afferenti la coesione sociale, Dickes et al. (2010), attraverso un'analisi fattoriale confermativa, identificano quattro fattori di primo ordine riferibili alla coesione: fiducia nelle istituzioni, solidarietà, partecipazione sociale e partecipazione politica.

Come abbiamo visto, ciascun autore ha fornito una differente concettualizzazione della coesione sociale e questo, come detto all'inizio, costituisce uno dei punti deboli dell'intero approccio accademico che ostacola lo sviluppo di una teoria della coesione sociale.

Vi è inoltre il limite riguardante il confine che separa le diverse discipline all'interno delle quali vengono proposte diverse formulazioni della coesione. La sociologia è maggiormente concentrata sull'evidenziare gli aspetti sistemici legati alla coesione sociale; la psicologia sociale risulta più interessata agli elementi collegati alla coesione di piccoli gruppi sociali.

In aggiunta, diversamente dall'approccio orientato alle politiche e con le eccezioni viste (soprattutto Chan et al. 2006), l'approccio accademico ha fornito poche definizioni operative utili a rilevare e misurare effettivamente il concetto al di là di aspetti meramente teorici.

1.3 Cos'è la coesione sociale? Una proposta

Abbiamo visto come sulla coesione sociale non vi sia pieno accordo, c'è chi la considera una proprietà della società considerata nel suo insieme, chi una proprietà riferibile ai gruppi. A nostro avviso, il concetto di coesione sociale può essere analizzato e considerato a più livelli di analisi.

Prendendo in considerazione tre domini di interesse – individui, gruppi ed istituzioni – possono essere identificati almeno sei livelli di analisi (tab.1.4): il primo si riferisce alle relazioni che sussistono fra individui; il secondo riguarda le relazioni che coinvolgono individui in rapporto ai gruppi, mentre il terzo livello si concentra sulle relazioni tra individui ed istituzioni. Il quarto e il quinto livello si riferiscono rispettivamente alle relazioni che intercorrono fra gruppi considerati nel loro insieme e relazioni che coinvolgono i gruppi e le istituzioni. Infine, il sesto livello riguarda le relazioni che sussistono fra diverse istituzioni.

Tab. 1.4 – I livelli della coesione sociale

Domini	Individui	Gruppi	Istituzioni
Individui	1	2	3
Gruppi		4	5
Istituzioni			6

Studiare le modalità attraverso le quali le istituzioni sono connesse chiama direttamente in causa un altro concetto ben noto alle scienze sociali: il concetto di integrazione sistemica, opposto a quello di integrazione sociale, si riferisce alle relazioni che sussistono fra le parti di un sistema sociale e i modi in cui tale istituzioni sono connesse fra di loro (Lockwood 1964).

In aggiunta ai livelli visti sopra, un ricercatore può anche investigare il livello relativo al grado di coesione che sussiste all'interno di un gruppo o all'interno di un'istituzione. Tali ulteriori livelli

implicano, rispettivamente, lo studio delle relazioni all'interno di un gruppo (ad esempio, meccanismi di identificazione col gruppo o discriminazione verso un *out-group*) e lo studio delle relazioni all'interno di un'istituzione (ad esempio, come le diverse parti di un'istituzione o di un'organizzazione sono connesse).

Seguendo la distinzione operata da Chan et al. (2006), si può ulteriormente studiare la tipologia di relazioni che sussistono fra i diversi domini (individui, gruppi, istituzioni) prendendo in considerazione la natura – orizzontale o verticale – delle relazioni implicate. Per esempio, prendendo in ipotesi il livello 6 presente in tab.1.4 che riguarda le relazioni che sussistono fra due o più istituzioni, si può indagare la dimensione orizzontale analizzando le relazioni che coinvolgono due o più istituzioni che giacciono sullo stesso piano (ad esempio, come due governi regionali di uno stato sono interconnessi) o la dimensione verticale esaminando le relazioni fra un'istituzione subordinata ed una posta su un piano superiore (ad esempio, le relazioni che intercorrono fra l'Unione Europea e i suoi Stati Membri).

A tale concettualizzazione si devono aggiungere ulteriori due componenti. Gli aspetti indagabili possono comprendere due ambiti fra loro distinti: si possono analizzare gli aspetti formali relativi agli atteggiamenti e alle rappresentazioni e gli aspetti sostanziali relativi ai comportamenti.

A complicare ulteriormente il quadro, si possono individuare facilmente differenti contesti (ambiti di analisi) nei quali la coesione può essere studiata. Un ricercatore potrebbe essere interessato a studiare la coesione sociale fra e all'interno di – per dirne alcuni – diversi quartieri, comunità, organizzazioni, associazioni, comuni, regioni, Paesi.

Seguendo un approccio relativo all'individualismo metodologico, nello studio che stiamo presentando, si prenderanno in considerazione soltanto i livelli che implicano la presenza di relazioni che coinvolgano l'individuo. Pertanto, la coesione sociale, sebbene sarà considerata una proprietà riferita alla società, sarà ascritta agli atteggiamenti, comportamenti e relazioni situate ad un livello individuale nei confronti di diverse sfere sociali.

1.3.1 Relazioni fra individui

Questo è il livello più basso al quale possiamo studiare la coesione sociale ed implica almeno la presenza di una relazione fra due individui. Vi sono tre possibili legami fra due soggetti: legami positivi, negativi e assenza di legame. Nel primo caso si fanno rientrare tutte le relazioni che generalmente vengono considerate positive (relazioni amicali, di coppia, familiari). Al contrario, le relazioni negative implicano una relazione ostile fra due attori. Mentre, l'assenza di legame denota la mancanza di ogni forma di relazione (non conoscenza, rottura di una precedente relazione). In aggiunta, una relazione oltre ad essere positiva, negativa o inesistente, può configurarsi come più o meno intensa. Le relazioni forti si riferiscono alle relazioni nelle quali le persone coinvolte mostrano un forte legame emotivo o hanno uno scambio di relazioni molto intenso. I legami deboli si riferiscono alle relazioni nelle quali gli individui interessati non sono in una condizione di vicinanza emotiva o sono scarsamente intense (per esempio le relazioni di vicinato). I legami deboli rivestono un'importanza pari a quella relativa ai legami forti. I legami deboli sono una sorgente fondamentale dalla quale ricavare informazioni e costituiscono un ponte essenziale che connette diverse reti relazionali (Granovetter 1973).

Le relazioni fra individui implicano il concetto di relazioni informali presenti sia nella sfera pubblica che privata. Lo studio dei legami informali si concentra su diversi tipi di relazioni. Innanzitutto, vi sono le relazioni che riguardano i rapporti familiari e parentali in generale. A tal proposito, per esempio, si può studiare la relazione che intercorre fra genitori e figli, se tale relazione è di supporto, mutuale o

svilupata e presente lungo tutto l'arco di vita del ragazzo. Anche i matrimoni, divorzi e relazioni di coppia sono legami fondamentali da investigare in quanto costituiscono importanti fattori che possono fornire supporto o generare ansie e frustrazioni. A queste, si aggiungono tutte le relazioni situate nella sfera pubblica. Uno dei legami più importanti è costituito dalle relazioni fra pari intese come rapporti amicali ma anche le relazioni che si generano in un una classe scolastica volte a verificare ad esempio se un bambino è in grado di stringere forti relazioni di amicizia o se esistono fenomeni di bullismo. Gli individui sviluppano relazioni anche all'interno dell'ambito lavorativo, nel quartiere in cui vivono e nel cosiddetto tempo libero legato alla pratica di qualche *hobby*. Allo stesso modo è fondamentale analizzare la natura delle relazioni informali che gli individui intrattengono, quanto spesso si incontrano con gli altri, che tipo e in che misura ricevono supporto, se questo supporto è reciproco o unidirezionale, in che misura le persone avvertono che gli altri siano disposti ad aiutarsi reciprocamente e quanto spesso partecipano ad attività sociali sia in termini assoluti sia quanto credono di partecipare in misura maggiore o minore comparato agli altri soggetti.

Tutti questi tipi di legami vanno a costituire il *network* sociale in cui un individuo è inserito. Inoltre, tenendo in debito conto il periodo storico in cui viviamo, non bisogna trascurare le relazioni che gli individui intrattengono *on-line*. Queste ultime stanno acquisendo sempre maggior rilievo nella definizione e composizione della rete sociale in cui un individuo è inserito.

1.3.2 Relazioni individuo-gruppo

I tipi di relazione appena illustrati sono riferiti all'ambito riguardante i legami informali all'interno dei gruppi primari. Le relazioni individuo-gruppo invece concernono i legami formali all'interno dei gruppi secondari. Questo specifico livello delle relazioni sociali è direttamente connesso al ruolo che ricopre ogni soggetto e al fatto di essere o meno associato o appartenente ad uno specifico gruppo o associazione. Tali organizzazioni possono essere politiche, culturali, associazioni sportive, comunità religiose, organizzazioni di volontariato e così via. In tali gruppi, i soggetti che ne fanno parte interagiscono fra di loro, si scambiano opinioni, rappresentazioni, idee e infine agiscono.

In questo caso, il ricercatore deve essere interessato al livello di partecipazione mostrato dai soggetti e valutarne in che misura ne sono coinvolti e la natura di tale coinvolgimento e delle attività svolte. Inoltre, risulta fondamentale indagare il tipo di legame che sussiste fra un individuo appartenente ad un gruppo o organizzazione e gli altri membri di altri gruppi sociali o altre associazioni così come valutarne il tipo di atteggiamento che i soggetti mostrano nei confronti di altri gruppi o organizzazioni. Per esempio, un ricercatore potrebbe essere interessato a studiare in che misura un soggetto ha un atteggiamento positivo nei confronti di gruppi appartenenti a minoranze etniche, comunità diverse o gruppi con un differente credo religioso. Allo stesso modo, in base ad una prospettiva sostanziale – comportamenti – potrebbe essere studiato quanto effettivamente un attore interagisce con gruppi diversi, che tipo di relazioni sussistono e se tali relazioni siano positive o negative.

Il livello di analisi riferito alla relazione individuo-gruppo costituisce una dimensione fondamentale in quanto coinvolge l'aspetto legato alla fratture sociali che sussistono spesso fra diversi gruppi etnici. In una società dove larghe masse di persone emigrano, dove persone con differenti stile di vita, religioni, *backgrounds* culturali ed etnie di origine diversa costantemente vivono a stretto contatto, è fondamentale che le relazioni inter-gruppo siano intense e positive. Diversamente, tale situazione potrebbe condurre a profonde fratture sociali che possono degenerare in fenomeni quali la segregazione, la ghettizzazione e varie forme di estremismo difficili da risanare. Le spaccature sociali

si verificano quando uno Stato è diviso in differenti gruppi sociali e i membri di tale gruppo intrattengono relazioni soltanto con gli altri membri del proprio gruppo e nessun legame viene instaurato con i soggetti appartenenti ad altri gruppi.

1.3.3 Relazioni individuo-stato

Per semplicità espositiva ad esemplare di istituzione è stato scelto lo Stato in quanto, attualmente, è la più importante istituzione politica e la cittadinanza costituisce ancora uno dei maggiori elementi di identificazione di un individuo.

Il livello riferito alle relazioni individuo-stato costituisce l'ultimo livello di analisi dei tre considerati qui. Tale dimensione risulta fondamentale al fine di investigare la coesione sociale in quanto scarsi livelli di interazione fra Stato e suoi cittadini e possibili elementi di tensione fra questi due attori pongono una seria minaccia alla coesione. Le relazioni positive fra Stato e suoi cittadini non costituiscono certamente un elemento che deve darsi per scontato. Vi sono numerosi esempi di Paesi in cui le relazioni fra Stato e cittadini non sono positive. I governi possono mettere in pratica azioni aggressive che possono variare da forme tenui di repressione, violazione dei diritti civili, fino ad arrivare a persecuzioni, omicidi, segregazione e genocidi. I cittadini, a loro volta, possono agire mettendo in pratica comportamenti aggressivi nei confronti del potere costituito. Le forme più lievi di protesta possono essere individuate nelle petizioni, nelle pubbliche manifestazioni, negli scioperi; le più aggressive si riferiscono a ribellioni, rivoluzioni, atti di terrorismo e via dicendo.

L'aspetto basilare delle relazioni che coinvolgono l'individuo e lo stato riguarda il concetto fondamentale di legittimità delle istituzioni. Le istituzioni, inquadrare all'interno di un contesto moderno e democratico, non sono, e non possono essere, basate sulla coercizione. Alla forza coercitiva si dovrebbe sostituire un altro aspetto: le istituzioni dovrebbero godere di alta reputazione – legittimità – nei confronti dei propri cittadini. Soltanto quando le istituzioni vengono percepite come legittime dagli attori sociali lo Stato può esercitare pienamente il suo ruolo.

La "qualità delle istituzioni", insieme alla legittimità, costituisce un ulteriore elemento essenziale al fine di evitare forme di dissoluzione sociale e quindi mantenere la coesione sociale.

Oltre a tale aspetto, la qualità delle istituzioni sotto forma di qualità dei servizi pubblici offerti (scuola, sanità, etc.) costituisce l'ambiente, o meglio la struttura – intesa come mezzi legittimi forniti per raggiungere legittimi obiettivi – in cui gli individui interagiscono e che costituisce l'orizzonte delle loro possibilità (Merton 1949).

1.3.4 Lo schema della coesione sociale

In definitiva la proposta concettuale, utilizzabile a fini operativi, di coesione sociale che avanziamo in questo contributo ha le seguenti caratteristiche:

- Partendo dalla definizione di Chan et al. (2006), intendiamo la coesione sociale come "una condizione riguardante le interazioni fra membri di una società, caratterizzati da una serie di atteggiamenti, norme e dalla loro manifestazione comportamentale che include la forza dei legami interpersonali nelle reti sociali primarie e secondarie, fiducia, partecipazione e riconoscimento di un insieme di regole, delle istituzioni democratiche e della società nel suo insieme"

- Tale definizione considera la coesione come una condizione e non come un processo, evitando di confondere i mezzi e i fini.
- Seguendo l'individualismo metodologico, la definizione si concentra sulle relazioni e i legami positivi che sussistono fra attori che agiscono all'interno di una società. Infine, nella definizione è espressamente riconosciuto che gli attori sociali operanti in una società devono riconoscerne l'insieme di norme, regole e la legittimità delle istituzioni democratiche. In tal modo si escludono tutte quelle forme che vedono gruppi molto coesi al loro interno ma indirizzati verso forme di comportamento criminale.
- Seguendo Whelan e Maître (2005), suggeriamo di analizzare la coesione sociale in base a tre ambiti di analisi: micro, meso e macro. Il primo ambito coinvolge relazioni fra individui, l'ambito meso riguarda le relazioni che sussistono fra individui e gruppi – come l'individuo si rapporta nei confronti di diversi gruppi – e, infine, il macro ambito prende in considerazione le relazioni fra individui e società.
- Ai tre ambiti di analisi possono essere aggiunti anche tre livelli di analisi, intesi come livelli gerarchici all'interno di una struttura multilivello. In questo caso, seguendo un approccio multilivello l'interesse si sposta sull'analisi delle relazioni che sussistono fra individui, gruppi ed aggregati di livello macro studiando le influenze che le unità di livello superiore esercitano sugli atteggiamenti e comportamenti individuali, i possibili effetti di interazione fra i livelli, distinguendone gli effetti in effetti individuali, di gruppo o contestuali.
- Inoltre, riprendendo Bollen e Hoyle (2001), verranno considerate due prospettive di analisi della coesione sociale all'interno dei tre livelli precedentemente illustrati: la prima prospettiva è quella soggettiva focalizzata sugli atteggiamenti degli individui, mentre la seconda – oggettiva – chiama in causa i comportamenti degli individui.

I tre livelli di analisi, gli ambiti e le due prospettive suggerite danno origine allo schema presentato in tab.1.5.

Tab.1.5 – Schema della coesione sociale

Ambiti di analisi	Prospettiva		Livelli di analisi
	soggettiva	oggettiva	
Micro	Relazioni fra individui in una prospettiva soggettiva	Relazioni fra individui in una prospettiva oggettiva	Micro L'unità di analisi è costituita esclusivamente dall'individuo. In questo caso l'approccio multilivello non viene applicato
Meso	Relazioni fra individui e gruppi in una prospettiva soggettiva	Relazioni fra individui e gruppi in una prospettiva oggettiva	Meso Un'unità di analisi di livello superiore in cui gli individui sono annidati gerarchicamente. Individuazione degli effetti di gruppo e le influenze che le unità di livello meso esercitano sugli individui
Macro	Relazioni fra individui e società in una prospettiva soggettiva	Relazioni fra individui e società in una prospettiva oggettiva	Macro Un'unità di analisi di livello superiore in cui gli individui sono annidati gerarchicamente. Individuazione degli effetti contestuali, ambientali e globali e le influenze che le unità di livello macro esercitano sugli individui

Lo schema concettuale proposto si pone l'obiettivo di coadiuvare ricercatori e studiosi nelle operazioni di misurazione della coesione sociale facilitando la selezione degli indicatori per ciascuno degli ambiti e delle prospettive individuate. Per esempio, considerando l'ambito di analisi micro, un indicatore soggettivo potrebbe riferirsi al grado di supporto che un individuo avverte di percepire dai soggetti a lui vicino; mentre un indicatore oggettivo potrebbe riferirsi alla frequenza con cui si incontra nel tempo libero con i propri amici.

Un ricercatore, per rendere più esplicito e chiaro l'intero percorso di ricerca, insieme alla concettualizzazione di coesione a cui fa riferimento, agli indicatori utilizzati per rilevarla, all'ambito spazio-temporale e all'unità di analisi dovrebbe dichiarare le dimensioni della coesione sociale, le prospettive e i livelli che prende in considerazione.

Come mostrato precedentemente, i livelli di analisi ai quali uno studioso può indagare la coesione sociale sono più numerosi di quelli presenti in tab.1.5 e possono includere lo studio della coesione sociale inter-gruppo o fra istituzioni considerate nel loro insieme. Nel nostro caso, seguendo la nostra prospettiva, ci siamo concentrati esclusivamente sulle relazioni che coinvolgono gli individui.

Lo schema non è soltanto utile allo studio della coesione sociale all'interno di una società, sebbene questa sia la prospettiva da noi selezionata. Un ricercatore, con l'aiuto dello schema proposto, può indagare la coesione che sussiste all'interno di un gruppo (coesione intra-gruppo), associazione, azienda, dipartimenti accademici, reparti di ospedali, organizzazioni, istituzioni e così via. Per esempio, potrebbe essere investigata la coesione sociale all'interno di un'azienda valutando le interazioni, gli atteggiamenti e i comportamenti dei dipendenti nei confronti degli altri colleghi a livello micro, verso il proprio gruppo di lavoro considerato nel suo insieme a livello meso e, infine, nei confronti dell'azienda valutata globalmente a livello macro.

In una fase successiva, seguendo una prospettiva guidata dall'approccio multilivello, un ricercatore può aggregare le variabili individuali o inserire variabili globali o contestuali al fine di indagare l'influenza delle unità di livello superiore sugli atteggiamenti e i comportamenti individuali studiandone inoltre le possibili forme di interazione fra i diversi livelli. All'approccio multilivello sarà dedicato il prossimo capitolo. Successivamente, nel terzo capitolo, tenendo ben presente lo schema teorico proposto si passerà a fornire la nostra proposta di operativizzazione del concetto di coesione sociale per poi valutarne empiricamente i possibili fattori che possono esercitare un'influenza sul concetto in esame.

2. L'analisi multilivello e i modelli ad equazioni strutturali multilivello

2.1 Introduzione

L'obiettivo di tale capitolo è presentare le principali tecniche di analisi utilizzate nella ricerca al fine di rendere maggiormente esplicite le implicazioni teoriche ed empiriche, la definizione operativa, il modello di misurazione della coesione sociale ed il modello ad equazioni strutturali multilivello della coesione sociale presentati nel prossimo capitolo.

Innanzitutto, si introdurrà l'analisi di regressione multilivello a cui sarà dato ampio spazio in quanto costituisce un tema poco esplorato in Italia. Successivamente, si presenteranno brevemente, in quanto ormai godono di una lunga tradizione, i modelli ad equazioni strutturali che saranno funzionali ad una piena comprensione della sezione successiva. Infine, si esporrà un nuovo approccio di analisi denominato modelli ad equazioni strutturali multilivello, quasi del tutto inesplorato in Italia che, come vedremo, ha il pregio di unire i benefici derivanti dall'analisi multilivello associandoli ai vantaggi offerti dai modelli ad equazioni strutturali classici.

2.2 L'analisi multilivello

Fin dalle origini delle scienze sociali, e della sociologia in particolare, gli studiosi si sono divisi sulla base dell'approccio ritenuto privilegiato nello studio del comportamento umano. Secondo la prospettiva olistica la società è intesa come una totalità con specifiche proprietà che le appartengono, non riducibili a quelle dei suoi membri. Al contrario, per l'approccio individualista, tutti i fenomeni sociali possono essere ridotti all'azione individuale. Nel primo caso, i fenomeni sociali hanno una propria specifica esistenza; nel secondo caso, quegli stessi fenomeni sono invece il prodotto dell'azione individuale.

Durkheim e Weber sono i rappresentanti principali di due tradizioni teoriche opposte: le teorie strutturali o sistemiche e le teorie dell'azione (Collins 1988).

Secondo Durkheim (1987), l'intenzionalità dell'individuo non è in nessun modo analizzabile in quanto non è né direttamente osservabile né attendibile al punto che le motivazioni di un'azione spesso sfuggono allo stesso soggetto agente. Per Durkheim le motivazioni degli individui non sono altro che un epifenomeno: le azioni individuali in realtà sono la causa di un fenomeno sociale. In particolare, il fatto sociale si configura come fenomeno che si differenzia da una semplice somma delle parti (individui) ma si pone con le sue caratteristiche riferite all'esteriorità rispetto all'individuo e alla forza coercitiva che esercita su esso. Weber ha una visione opposta ritenendo come la sociologia debba avere come oggetto di analisi l'individuo e il suo agire (1922). L'autore individua quattro tipi di agire sociale: 1) l'agire razionale rispetto allo scopo; 2) l'agire razionale rispetto al valore; 3) l'agire affettivo; 4) l'agire tradizionale. I tipi di azione sociale non sono né esaurienti dei modi dell'agire umano né sono direttamente osservabili nel comportamento concreto, piuttosto rappresentano dei tipi puri – costruzioni sociologiche – di scopo eminentemente euristico (Weber 1922). L'agire razionale, per le sue caratteristiche di intelligibilità, rappresenta l'oggetto privilegiato per l'analisi sociologica. Gli aspetti sociali di carattere collettivo, invece, sono dovuti ai singoli individui che indirizzano il proprio agire in base alle rappresentazioni che si fanno di tale collettività.

L'analisi multilivello permette di superare la tensione fra una prospettiva microsociologica ed una macrosociologica. Infatti, l'analisi multilivello permette di connettere gli individui alla società e ai

contesti in cui agiscono. Tale approccio di analisi, infatti, rende esplicito il fatto che fattori di livello superiore (ossia che si trovano ad un livello epistemologico al di sopra di quello individuale) possono esercitare un'influenza diretta, o indiretta, sul comportamento individuale. I livelli di analisi non devono essere pensati in nessun modo come "livelli ontologici", piuttosto come livelli epistemologici, utili alle procedure di indagine. Tanto è vero che, contrariamente al senso comune, il livello di analisi individuale non necessariamente costituisce il livello base, ma possono essere individuati livelli sub-individuali – per esempio le diverse occorrenze o punteggi su scale di atteggiamento forniti da un individuo in un'indagine longitudinale. Inoltre, i livelli possono variare a seconda delle discipline prese in considerazione: l'insegnante, la classe e la scuola negli studi sull'efficienza del sistema educativo, la regione o lo Stato in sociologia, il reparto e l'ospedale negli studi sull'organizzazione aziendale.

Al fine di studiare il comportamento dell'individuo adottando una prospettiva globale e non parziale, non si può in nessun modo ignorare l'influenza che un gruppo sociale (o qualsiasi altra forma di aggregato) esercita sui comportamenti ed atteggiamenti dei suoi membri e, di rimando, l'influenza che il singolo membro esercita sull'intero gruppo.

Ignorare tale effetto, da un lato, potrebbe condurre ad un'analisi incompleta e statisticamente errata (individuando relazioni che in realtà non sussistono o viceversa affermarne erroneamente l'assenza) dall'altro lato, a livello teorico, si perde una grande occasione nel cercare di spiegare i fenomeni sociali solo attraverso il ricorso a proprietà individuali escludendo il contributo di proprietà contestuali e strutturali dal modello teorico sviluppato. In altre parole, si rinuncia a priori a studiare l'effetto del contesto sui comportamenti ed atteggiamenti individuali nonché l'interazione fra ambiente e individuo.

Nella prospettiva multilivello, così come noi la intendiamo, la società (il contesto, l'aggregato, etc.) non è vista come un'entità che esercita la sua influenza indipendentemente dal comportamento individuale, dalle ragioni e dalle motivazioni che guidano il singolo individuo – così avviene per esempio nella prospettiva olistica durkheimiana. Piuttosto, la struttura sociale viene intesa come sistema di possibilità e costrizioni. In tal modo, gli individui sono il risultato dell'azione congiunta esercitata dal proprio stesso sistema di credenze/opinioni/atteggiamenti e dalla struttura delle possibilità/costrizioni del sistema sociale (Giddens 1984).

L'approccio multilivello, permettendo esplicitamente di considerare comportamenti ed atteggiamenti una funzione non solo delle caratteristiche individuali ma anche delle costrizioni o condizionamenti dell'ambiente (la maggioranza dei quali esercitano i loro effetti sull'individuo senza che costui ne abbia consapevolezza), consente anche di esaminare i cosiddetti effetti perversi dell'azione sociale (Boudon 1977).

In tale prospettiva, il modello prevede, inoltre, una pluralità di livelli – teoricamente infinita ma praticamente limitata da numerosi fattori operativi come il disegno campionario, l'eccessiva complessità statistica, la difficoltà di reperire i dati a livelli superiori a quello individuale – che, come detto, possono variare da disciplina a disciplina e in base all'obiettivo cognitivo prefissato. Non c'è un limite al numero di livelli individuabili da un ricercatore, per esempio gli individui appartenenti a specifici quartieri, contenuti in città, province, regioni, Stati e, infine, organi sovra-nazionali. Ma in uno studio sulla partecipazione elettorale, per esempio, il livello riferito ai quartieri o alle città potrebbe perdere di rilevanza a favore delle circoscrizioni elettorali. La pluralità di livelli e il suo variare in base agli obiettivi prefissati è fondamentale anche perché permette di liberarci della prospettiva che pone i soggetti in opposizione alla società e viceversa. In altre parole, la contrapposizione olistica vs. individualismo perde di pregnanza a favore di un approccio volto ad esplorare i diversi livelli in cui un individuo è inserito e l'interazione fra tali livelli e le proprietà individuali.

L'analisi multilivello, permettendo di rilevare l'influenza che le caratteristiche di un ambiente e/o dei gruppi sociali esercitano sul comportamento e gli atteggiamenti individuali, ci consente in tal modo di evitare di incorrere sia nella fallacia ecologica sia nel suo inverso – la fallacia atomistica (Lazarsfeld e Menzel 1961). Nel primo caso, disporre dei dati raccolti a livello individuale (ossia delle distribuzioni congiunte e non solo marginali) scongiura il rischio di fallacia ecologica; nel secondo caso, la fallacia atomistica è eliminata inserendo gli individui nei contesti in cui agiscono.

Inoltre, i contesti individuabili, ovvero i livelli dell'analisi multilivello, non necessariamente devono coincidere con spazi fisici geograficamente delimitati. Tali contesti possono essere anche delimitati socialmente, come le reti relazionali in cui gli individui sono inseriti, gli ambienti lavorativi, gli status socio-economici a cui appartengono.

Il modello in fig.2.1 illustra come l'approccio multilivello può connettere il livello micro e il livello macro d'analisi. Sono mostrati tre diversi modelli fra loro indipendenti. Il primo modello, quello contrassegnato dalla linea superiore non tratteggiata, rappresenta il modello aggregato, ovvero quando si fa ricorso ad indicatori che sono il frutto di un'aggregazione, come generalmente avviene nelle indagini ecologiche che fanno ricorso agli indicatori sociali. Un esempio di tale modello potrebbe essere il tasso di natalità di uno Stato fatto regredire sulla percentuale di immigrati di quel Paese. Al contrario, la linea inferiore non tratteggiata rappresenta il modello individuale – ossia il tipo di modello maggiormente diffuso fra i ricercatori. Tale modello si riferisce generalmente alla spiegazione di una proprietà individuale attraverso diverse proprietà esplicative individuali (per esempio il livello di felicità di un individuo fatto regredire su una serie di predittori come l'età, il genere, il livello di istruzione, il reddito, lo stato civile, etc.).

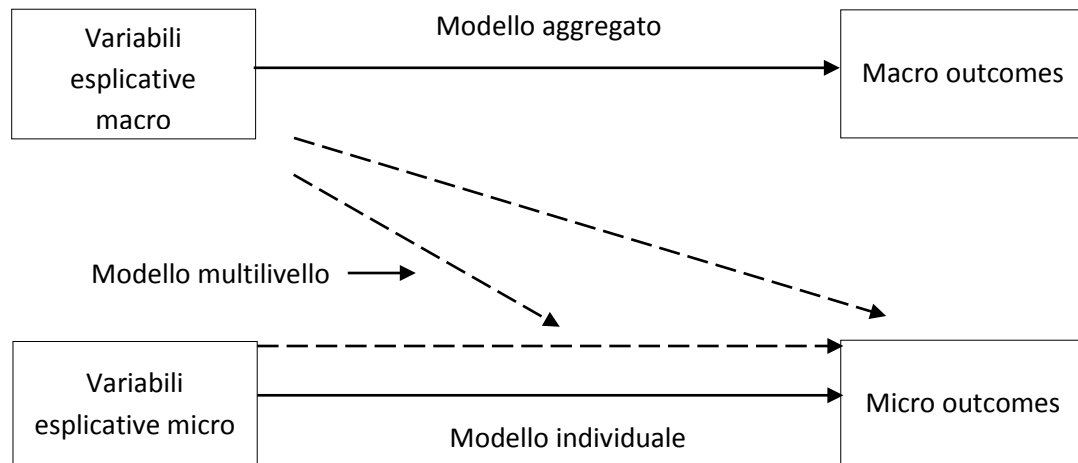
Le linee tratteggiate, invece, mostrano concettualmente il modello multilivello base e come tale modello riesca a connettere i livelli micro e macro.

Supponiamo che un ricercatore sia interessato a studiare i fattori che influenzano il risultato ad un test che rileva la conoscenza matematica (micro *outcomes*) somministrato ad un campione di studenti delle scuole superiori appartenenti (*nested*) a diverse scuole del territorio italiano. Ad un livello individuale un possibile *set* di variabili esplicative potrebbe essere composto dal "tempo totale speso per lo studio della matematica", "il livello di scolarizzazione dei genitori dello studente", "lo status socio-economico" e così via. Tuttavia, tale modello non prende in considerazione altre variabili collocate ad un livello superiore a quello individuale e pertanto non direttamente collegate all'individuo ma che ugualmente possono influenzarne la relazione. Per esempio, la variabile "anni totali di esperienza professionale del docente" potrebbe essere un valido fattore esplicativo di livello macro, così come la presenza in alcune scuole di specifici programmi governativi volti a migliorare la conoscenza generale della matematica da parte degli studenti. Quelle appena illustrate sono le cosiddette variabili contestuali che assumono lo stesso valore per ogni soggetto intervistato appartenente allo stesso contesto aggregativo (nel nostro caso le scuole). Il modello, però, può prevedere anche l'inserimento delle cosiddette variabili strutturali che sono il frutto di un processo di aggregazione; un esempio relativo al nostro caso potrebbe riguardare il livello medio di conoscenza della matematica in una specifica scuola che potrebbe esercitare un effetto sulla conoscenza della matematica del singolo studente (in una classe in cui il livello di conoscenza della matematica è generalmente alto uno studente può essere invogliato a studiare maggiormente).

Le variabili esplicative macro non solo possono esercitare un effetto diretto sulla variabile dipendente a livello micro, ma possono anche interagire con le variabili esplicative micro influenzandone la relazione con la variabile dipendente.

Per esempio, la variabile “anni di esperienza dell’insegnante” può influenzare la relazione a livello individuale fra il totale di ore spese nello studiare matematica da parte di uno studente (insegnanti con maggiore esperienza possono incentivare maggiormente lo studio della matematica attraverso un meccanismo che coinvolge la maggiore efficacia dell’insegnamento) e il risultato al test di matematica.

Fig.2.1 – Modello aggregato, individuale e multilivello



Nell'esempio precedente si è fatto riferimento ad un campione di studenti delle scuole superiori appartenenti (*nested*) a diverse scuole. Non è un aspetto irrilevante. L'analisi multilivello assume senso soltanto se le unità di analisi oggetto della nostra indagine sono gerarchicamente organizzate, ovvero contenute l'una dentro l'altra. È il caso, per esempio, degli studenti inseriti in classi, appartenenti a loro volta a scuole, province, regioni, Stati.

Riassumendo, possiamo individuare almeno due caratteristiche specifiche dei modelli multilivello: nella sua forma minima il modello multilivello deve prevedere almeno due livelli di analisi; tali livelli devono essere ordinati gerarchicamente (van de Vijver, van Helmer e Poortinga 2008).

L'esistenza di più livelli di analisi chiama in causa il concetto di isomorfismo. Particolarmente rilevante nelle indagini comparative transnazionali (*cross-cultural*), tale concetto si riferisce al parallelismo nella struttura e nei costrutti fra diversi livelli di analisi: indica che la struttura dei costrutti è simile o equivalente fra i differenti livelli di analisi. In altre parole, l'interpretazione e il contenuto semantico dei costrutti sono uguali fra i diversi livelli di analisi considerati. Empiricamente, un modello è isomorfo quando lo stesso modello di misurazione (per i modelli di misurazioni multilivello si rimanda al paragrafo 2.4) si adatta ai dati sia a livello individuale che a livello aggregato. In termini più precisi significa che lo stesso costrutto latente può essere usato sia a livello individuale che aggregato, che tale costrutto ha un'associazione con uno specifico indicatore A a livello individuale, infine, che tale costrutto latente ha un'associazione con uno specifico indicatore A a livello aggregato. Quando vi è isomorfismo, l'associazione fra i diversi indicatori a livello individuale ed aggregato mostrano la stessa struttura interna e lo stesso costrutto può essere utilizzato per spiegare la varianza sia a livello individuale che aggregato (Fontaine 2008). I modelli non isomorfi, al contrario, non mostrano questa equivalenza fra i diversi livelli di analisi. Un modello è non isomorfo quando, sebbene troviamo la stessa struttura fattoriale in ciascuno dei sotto gruppi, questa differisce passando dal livello individuale al livello aggregato. Le relazioni fra gli indicatori sono diverse passando da un livello ad un altro. Uno degli esempi più calzanti a riguardo si riferisce al concetto di individualismo/collettivismo individuato da Hofstede (1980). A livello macro – dello Stato – l'analisi empirica del concetto mostra la presenza di un

solo fattore –il *continuum* individualismo/collettivismo – ma a livello micro – dell'individuo – tale struttura fattoriale non si manifesta richiedendo il ricorso a due fattori distinti. Secondo Triandis (2001), l'individualismo e il collettivismo sono due opposti poli di una singola dimensione a livello di Stato ma a livello individuale il concetto si sdoppia richiedendo il ricorso a due costrutti che l'autore definisce *idiocentrismo* e *allocentrismo*. Tale modello, quindi, con un costrutto a livello di Stato e due a livello individuale può dirsi non isomorfo in quanto sul piano operativo i concetti, passando da un livello ad un altro, risultano connessi con indicatori diversi e dunque acquistano un diverso contenuto semantico.

Un'altra caratteristica contraddistingue i modelli multilivello, ossia la presenza di concetti e misure definibili *intrinseche* e *derivate*. Le proprietà intrinseche, dopo essere state operativizzate, sono variabili che vengono usate al loro livello "naturale", o meglio utilizzate allo stesso livello dell'unità di raccolta, per esempio il livello di tolleranza verso gli immigrati di un individuo spiegato attraverso il titolo di studio degli intervistati oppure la qualità del sistema educativo di uno Stato spiegato attraverso la spesa in percentuale del PIL per l'educazione. Le variabili derivate, invece, sono raccolte ad uno specifico livello ma utilizzate ad un altro attraverso processi di aggregazione o disaggregazione. Per esempio il livello medio di tolleranza in una classe scolastica – ottenuta per aggregazione dei punteggi individuali – usata come variabile esplicativa della tolleranza di un singolo alunno in una specifica classe. La suddivisione in variabili intrinseche e derivate non è un esercizio analitico, ma assume rilevanza in quanto ignorare tale distinzione può causare confusione concettuale. Infatti, i processi di aggregazione o disaggregazione di una variabile non si risolvono meramente in trasformazioni statistiche ma possono condurre ad una vera e propria modifica del contenuto semantico di una variabile. Per esempio, la variabile genere a livello individuale è una variabile dicotomica che dà conto delle differenze fra uomo e donna, ma a livello aggregato diventa continua e si riferisce alla proporzione di maschi e femmine sull'intera popolazione.

Da un punto di vista teorico, fondamentale è la scelta del livello appropriato di analisi valutando a quale livello sia più opportuno concettualizzare il costrutto oggetto di analisi, se a livello individuale o a livello collettivo/aggregato (Fischer 2008). Un errore di concettualizzazione potrebbe portare a gravi implicazioni per la teoria, operativizzazione e misurazione del concetto così come per l'analisi e l'interpretazione dei dati. Passare da un livello ad un altro di analisi (*shifting levels*) può condurre a numerosi problemi di interpretazione. In particolare, il cosiddetto "errore strutturale" si riferisce all'attribuzione di un concetto ad un livello al quale non appartiene, mentre "l'errore di livello" (*level error*) si riferisce all'incorretta attribuzione di valori – punteggi – da un livello di aggregazione ad un altro. A ciò si aggiungono altri due diversi tipi di errore: di aggregazione e di disaggregazione. Il primo si verifica quando caratteristiche appartenenti ad un livello gerarchico inferiore (es. individui) vengono attribuite ad un livello gerarchico superiore (es. organizzazioni, Stati, etc.). L'errore di disaggregazione, invece, riguarda l'attribuzione di caratteristiche appartenenti ad un livello gerarchico superiore ad un livello gerarchico inferiore (Van de Vijver e Poortinga, 2002). L'incrocio dei quattro errori conduce a quattro possibili combinazioni (tab.2.1): l'errore 1 (disaggregazione - strutturale) si verifica quando si assume che un costrutto misurato ad un livello gerarchico superiore abbia lo stesso significato ad un livello gerarchico inferiore; l'errore 2 (aggregazione - strutturale) è l'opposto del precedente verificandosi quando si ipotizza che un costrutto misurato ad un livello gerarchico inferiore abbia lo stesso significato a un livello superiore; l'errore 3 (aggregazione - livello) riguarda l'assegnazione di punteggi raccolti ad un livello gerarchico inferiore ad unità appartenenti ad un livello gerarchico superiore; infine, l'errore 4 (disaggregazione - livello) si manifesta con l'attribuzione di punteggi e

valori ottenuti su unità gerarchicamente superiori ad unità di livello inferiore (Van de Vijver e Poortinga 2002).

Tab.2.1 – Tipologia di errori “shifting levels”

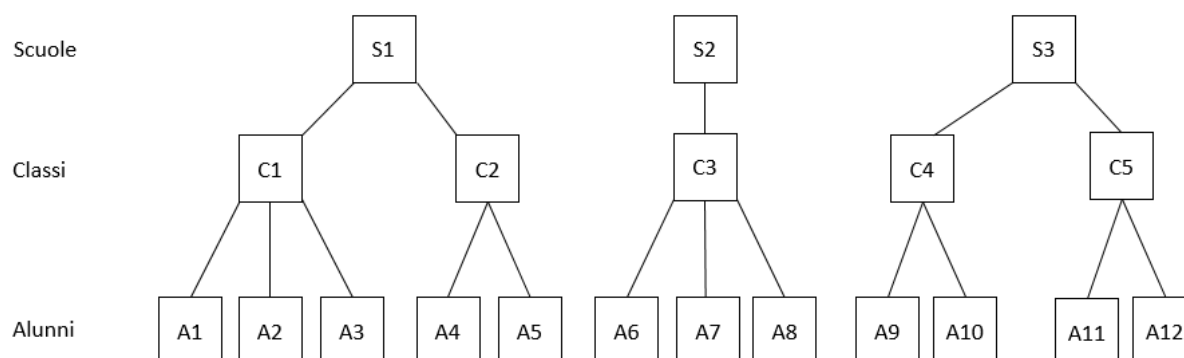
	Errore strutturale	Errore di livello
Errore di disaggregazione	1	4
Errore di aggregazione	2	3

In riferimento all'attribuzione di un concetto ad un livello o ad un altro, due tipi di meccanismi teorici sono ipotizzabili: processi (o teorie) *top-down* e processi (o teorie) *bottom-up*. Le teorie *top-down* si concentrano sull'influenza che il contesto esercita sugli atteggiamenti e comportamenti individuali. In questo caso, le variabili di secondo livello (o maggiore) possono esercitare un effetto diretto sulle unità di primo livello o possono moderare la relazione fra variabili di primo livello. Diversamente, le teorie *bottom-up* postulano che i fenomeni collettivi possono avere la loro origine dall'azione esercitata dagli individui attraverso un processo di interazione che genera un costrutto emergente. A tal proposito è importante distinguere fra costrutti emergenti, collettivi ed aggregati (Fischer 2008). I costrutti collettivi risiedono al livello collettivo di analisi e rappresentano “descrizioni” di fenomeni collettivi in cui viene postulato un processo *top-down*; un costrutto emergente è sempre un costrutto che risiede a livello collettivo ma è generato dall'interazione fra unità di livello inferiore; un aggregato, invece, è sempre costituito dalle unità di livello inferiore ma non rappresenta nessun tipo di costrutto collettivo riflettendo semplicemente l'aggregazione di punteggi individuali (si pensi ad esempio al livello medio di conoscenza della matematica in una classe scolastica costruito tramite aggregazione dei singoli punteggi degli studenti).

2.2.1 I presupposti dell'analisi multilivello

Si è detto in precedenza che applicare un approccio multilivello ha senso soltanto in presenza di unità di analisi “annidate gerarchicamente” (*nested*). Con tale espressione s'intende una struttura in cui ogni unità h è contenuta in un gruppo che costituisce l'unità di livello $h+1$ (fig.2.2). In altre parole, si è in presenza di una struttura gerarchica quando un individuo è “contenuto” all'interno di un contesto d'analisi più ampio che lo contiene, come in fig.2.2 con gli alunni inseriti in classi appartenenti a scuole.

Fig.2.2 – Esempio di struttura gerarchica dei dati con tre livelli



L'analisi multilivello permette di analizzare le relazioni fra variabili rilevate a differenti livelli di analisi simultaneamente. La struttura ipotizzata è quindi una struttura gerarchica con gli individui annidati (*nested*) all'interno di gruppi sociali e questi innestati all'interno di specifici contesti/società³. Una tale struttura produce un campione che a sua volta è ordinato gerarchicamente in quanto ovviamente riproduce in piccolo la struttura della popolazione.

Analizzare dei dati ordinati gerarchicamente con tecniche di analisi che non permettono di tener conto simultaneamente di più livelli conduce a tre problemi interconnessi. Il primo è di carattere teorico. L'annidamento gerarchico (*nesting*) non è da intendersi soltanto come un problema metodologico da trattare con le dovute tecniche di analisi ma rappresenta una ricca opportunità analitica per il ricercatore. Ignorando la struttura gerarchica dei dati si perde la possibilità di tener conto degli effetti che il contesto esercita sul comportamento e sugli atteggiamenti dell'attore sociale. Attraverso l'approccio multilivello, i modelli di spiegazione del comportamento umano si arricchiscono di variabili che trascendono il livello individuale permettendo al ricercatore di tendere verso modelli correttamente specificati. Il problema del *nesting* non è certo nuovo per le scienze sociali e non nasce certo con l'analisi multilivello. Un esempio a riguardo è costituito dall'indagine di Barr e Dreeben (1983) dove il quoziente intellettivo viene trattato sia come caratteristica dei singoli studenti (livello individuale) sia come caratteristica del gruppo in cui gli studenti sono inseriti (livello aggregato). L'indagine, a dispetto delle intenzioni degli autori, è essenzialmente un'analisi mono-livello in quanto sia il QI individuale che il QI aggregato vengono analizzati entrambi come caratteristica individuale. Il *nesting* è soltanto formalmente riconosciuto attraverso il calcolo delle medie di gruppo del QI ma viene sostanzialmente ignorato in quanto le due variabili (QI individuale e QI di gruppo), inserite nella stessa equazione attraverso il metodo dei minimi quadrati, vengono di fatto analizzate come se fossero entrambe proprietà individuali. Sotto questo aspetto, come si diceva, l'analisi multilivello offre nuove opportunità analitiche ampliando le possibilità del ricercatore e permettendo a quest'ultimo di tener conto degli effetti contestuali e strutturali esercitati sugli atteggiamenti/comportamenti dell'individuo. Il secondo problema è di carattere metodologico. Analizzare con tecniche statistiche classiche dati organizzati gerarchicamente viola numerosi assunti dell'analisi di regressione. In particolare, viene violato l'assunto della "indipendenza delle osservazioni". Nei campioni con una struttura di dati annidati gerarchicamente, le osservazioni in genere non sono indipendenti. Per esempio, gli studenti di una classe tenderanno ad essere più simili fra di loro rispetto ad altri di un'altra classe a causa, per esempio, di processi di selezione (alcune scuole tenderanno ad attrarre studenti con uno status socio-economico maggiore rispetto ad altre che possono selezionare studenti con uno status socio-economico più basso) o a causa del "vissuto" comune che gli studenti di una stessa classe condividono (Hox 2010). Quando le osservazioni non sono indipendenti, i residui prodotti dall'equazione di regressione mostreranno *autocorrelazione* violando un ulteriore assunto ossia quello della "indipendenza degli errori". Tutto ciò produrrà stime errate degli "errori standard" inficiando così i test di significatività. L'annidamento gerarchico dei dati, inoltre, produce il cosiddetto "effetto del disegno campionario" (*design effect*). In sostanza, il campione effettivo è più piccolo di quello che abbiamo effettivamente raccolto: abbiamo meno informazioni di quante in realtà ne abbiamo raccolte. Il *design*

³ Per semplicità di esposizione ci riferiamo ad individui, gruppi, società, ma, come detto in precedenza, non c'è un limite teorico all'individuazione di livelli intermedi. Inoltre, non necessariamente l'individuo costituisce l'unità di livello inferiore: negli studi longitudinali, le misure ripetute sono le unità di base innestate nelle unità di secondo livello costituito dagli individui. Infine, i livelli possono riferirsi anche ad altri tipi di aggregati e non necessariamente implicare l'individuo come unità di base: per esempio i reparti di un ospedale (unità di primo livello) annidati in ospedali (livello-2), annidati in distretti sanitari (livello-3) e così via.

effect è il prodotto del numero medio delle osservazioni in ogni gruppo e del “coefficiente di correlazione intraclasse” – *intraclass correlation coefficient* (ICC) – (che è una misura della non indipendenza fra osservazioni, ovvero della omogeneità all'interno di un gruppo dato dal rapporto fra la varianza *between*, ossia fra i gruppi, e la varianza totale) (Hox 2010). A titolo di esempio, si consideri un campione abbastanza ampio composto in totale da mille studenti con un numero di osservazioni medio per classe di 25 soggetti ed un ICC di 0.1 (abbastanza comune nelle scienze sociali): il campione effettivo risulta essere di 294 a fronte dei mille originali⁴. Maggiore saranno le osservazioni annidate gerarchicamente in gruppi, più grandi saranno i costi in termini di campione effettivo. Tale effetto è esacerbato quando il disegno campionario prevede un numero relativamente basso di gruppi associato ad un ampio numero di casi presenti in ogni *cluster* (Hox 2010; Snijder e Bosker 1999). Nell'esempio precedente se ipotizziamo un numero di osservazioni medio per gruppo di 100 soggetti, il campione effettivo scende a 91,7 a fronte dei mille originali.

Un altro assunto violato in presenza di dati ordinati gerarchicamente riguarda l'invarianza dei coefficienti di regressione e dell'intercetta. In un'analisi di regressione lineare classica si assume che i coefficienti (coefficiente angolare della retta o del piano ed intercetta) siano gli stessi per tutti i sotto-gruppi presenti nella popolazione. In realtà, a seconda dei gruppi analizzati i coefficienti potrebbero mostrare valori differenti indicando la presenza di relazioni diverse (a volte anche di segno opposto) fra la variabile dipendente e i regressori all'interno dei diversi sotto-gruppi. L'analisi multilivello permette di accomodare tale eventualità permettendo ai diversi coefficienti di variare fra i diversi *cluster* presenti nel nostro campione.

Oltre a problemi di carattere teorico e metodologico, se ne può individuare anche uno empirico. Come detto, in presenza di dati annidati gerarchicamente le osservazioni non sono indipendenti, ciò influisce sulla stima degli errori standard il cui valore risulterà, in genere, deflazionato, inficiando i test di significatività. Da un punto di vista sostantivo, ciò significa che un ricercatore può individuare relazioni significative fra variabili che in realtà non sussistono, traendone conclusioni teoriche errate. In altre parole, aumentano le probabilità di rifiutare erroneamente l'ipotesi nulla.

L'analisi di regressione multilivello non è soltanto una tecnica che consente di analizzare strutture di dati annidati gerarchicamente laddove la regressione classica non può più essere utilizzata, ma si pone anche finalità sostantive permettendo, come detto, di misurare⁵ gli effetti contestuali sui comportamenti/atteggiamenti individuali. A tal proposito, si possono distinguere almeno tre tipi di effetti contestuali (Stanat e Ludtke 2008). Il primo è il cosiddetto *direct cross-level effect*, in cui la variabile contestuale Z esercita un effetto diretto sulla variabile dipendente Y controllando al contempo l'effetto esercitato dalla variabile di primo livello X (fig.2.3). Un esempio potrebbe riguardare la percentuale di PIL destinata all'istruzione (Z) e quanto questa incida sui risultati scolastici (Y) tenendo costante il supporto fornito dai genitori (X) e assumendo che quest'ultimo sia costante nei diversi Stati oggetto di indagine. Tali modelli vengono denominati “modelli ad intercetta variabile” (*random-intercept model*). Nel *cross-level effect modification*, invece, il fattore contestuale non influenza più la variabile dipendente ma l'intera relazione fra questa e la variabile indipendente

⁴ La formula per calcolare il *design effect* è la seguente: $design\ effect = 1 + (n - 1) * ICC$. Il campione effettivo non è altro che il campione originario diviso il *design effect*.

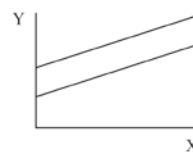
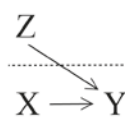
⁵ Il termine misurazione è usato qui in senso “ingenuo”, per antonomasia, indicando tutte le operazioni di traduzione empirica e rilevazione di una proprietà senza nessuna pretesa e riferimento all'atto di misurazione vero e proprio. Si è consapevoli, infatti, che nelle scienze sociali non si può parlare di vera e propria misurazione mancando i requisiti fondamentali fra i quali assenza di un'unità di misura replicabile ed intersoggettiva, necessaria cooperazione dei soggetti il cui stato è rilevato (Marradi, 1981).

(fig.2.3). In pratica la variabile contestuale Z funge da moderatore della relazione individuale fra X e Y. Tale effetto nella letteratura multilivello prende il nome di *cross-level interaction*. Tornando all'esempio precedente, la relazione fra supporto fornito dai genitori (X) e risultati scolastici (Y) potrebbe variare in funzione della percentuale di PIL destinata all'istruzione (Z): la relazione potrebbe essere più forte nei Paesi con una percentuale di PIL destinata all'istruzione e relativamente più bassa rispetto ai Paesi dove la spesa per l'istruzione in rapporto al PIL è maggiore. Il terzo tipo di effetto contestuale viene definito *indirect cross-level effect*. In questo caso, l'effetto esercitato dal fattore contestuale sulla variabile dipendente è mediato da una variabile mediatrice o moderatrice posta a livello individuale (fig.2.3). Nel caso della presenza del mediatore, ritornando sempre al nostro esempio, la relazione fra percentuale di PIL destinata all'istruzione (Z) e risultati scolastici (Y) potrebbe essere mediata dal capitale culturale di cui lo studente dispone (M). In presenza di un moderatore invece, l'effetto esercitato dalla variabile contestuale percentuale di PIL destinata all'istruzione (Z) sulla relazione fra supporto fornito dai genitori (X) e risultati scolastici (Y) è mediato da una variabile posta a livello individuale come, per esempio, il supporto personalizzato che lo studente riceve a scuola nelle ore extrascolastiche (M).

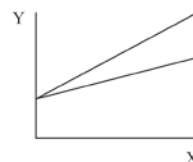
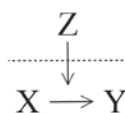
Nel prossimo paragrafo cercheremo di formalizzare tali modelli. Prima però presenteremo il modello base di regressione multilivello.

Fig.2.3 – Gli effetti contestuali nell'analisi multilivello

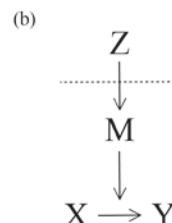
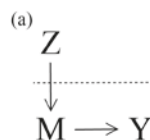
(1) Direct cross-level effect



(2) Cross-level effect modification



(3) Indirect cross-level effect



2.2.2 Il modello di regressione multilivello

I modelli multilivello, così come i classici modelli di regressione lineare, si pongono come strumento utile a spiegare o predire i valori di una variabile dipendente attraverso una serie di variabili indipendenti o regressori. A differenza dei modelli di regressione classici, la particolarità risiede nel fatto che l'equazione di regressione tiene conto della struttura gerarchica dei dati cercando in tal modo di integrare l'analisi a livello individuale (micro) e l'analisi a livello aggregato (macro). Quindi, a differenza della regressione classica che assume l'esistenza di un'unica equazione valida per tutti i gruppi in cui la popolazione è suddivisa, i modelli multilivello prevedono non una singola equazione ma un sistema di equazioni gerarchiche che assumono diversi parametri a seconda del gruppo considerato. Inoltre, le variabili esplicative non devono riferirsi esclusivamente al livello individuale, ma possono provenire da tutti i diversi livelli della struttura gerarchica presi in considerazione nel campione.

Quindi, affinché si possa applicare l'approccio multilivello all'analisi dei dati devono essere soddisfatte almeno due condizioni: che il campione sia strutturato gerarchicamente (*nested data*) e che siano presenti variabili rilevate su diversi livelli di tale struttura. Se una delle due condizioni non è soddisfatta non ha senso applicare la tecnica multilivello, risultando sufficiente il ricorso alla regressione classica. Come esempio di struttura gerarchica di dati considereremo un campione di studenti (n_j) appartenenti a diverse classi (J). Il campione sarà così composto da due livelli: quello degli studenti, e quello delle classi. L'obiettivo sarà quello di dar conto della variazione della variabile indipendente individuando i fattori che possono influenzarla ricorrendo a variabili esplicative rilevate ad entrambi i livelli della gerarchia. Al primo livello, abbiamo la variabile indipendente costituita dal risultato ad un compito di comprensione del testo; sempre al primo livello abbiamo due variabili indipendenti: il genere degli studenti (codificato 0 = ragazzi, 1 = ragazze); il capitale culturale (rilevato come media degli anni totali di studio dei genitori). Al secondo livello, abbiamo la variabile esplicativa contestuale riferita agli "anni di esperienza dell'insegnante".

L'analisi si compone di due fasi concettualmente separate (si veda anche Hox 2010; Snijder e Bosker 1999). Nella prima fase, vengono realizzati differenti modelli di regressione in ciascuna classe al fine di predire la variabile risposta Y attraverso le variabili esplicative X :

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \varepsilon_{ij} \quad (1)$$

Nell'equazione 1 β_{0j} è l'intercetta, β_{1j} e β_{2j} sono i due coefficienti di regressione per le due variabili indipendenti rilevate a livello-1, mentre ε_{ij} costituisce il termine di errore, o residuo. Rispetto al modello di regressione classico, qui si assume che ciascuna classe abbia una differente intercetta β_{0j} e diversi coefficienti di regressione β_{1j} e β_{2j} . Nell'equazione tale aspetto è rimarcato dall'aggiunta del pedice J alle notazioni dei diversi coefficienti, indicando che ogni classe J ha il proprio coefficiente. Tali coefficienti, variando in funzione della classi (dei gruppi di livello-2), vengono denominati "coefficienti casuali" (*random coefficients*). Si assume che tali coefficienti siano normalmente distribuiti fra le classi. Nella seconda fase, dal momento che i coefficienti β_j variano in ciascuna classe, l'obiettivo è spiegare tale variazione attraverso le variabili esplicative rilevate a livello-2. Le equazioni di secondo livello saranno:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j} \quad (2)$$

e

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j} \quad (3)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}Z_j + u_{2j} \quad (4)$$

L'equazione 2 predice l'intercetta casuale β_{0j} attraverso la variabile esplicativa di secondo livello Z. In altre parole, predice il risultato medio al compito di comprensione del testo in una classe attraverso la variabile "anni di esperienza dell'insegnante". L'interpretazione delle equazioni 3 e 4 è più complessa: la prima afferma che la relazione fra la variabile Y e la variabile X_1 (genere) dipende dalla variabile Z, ossia il numero di anni di esperienza dell'insegnante; la seconda, allo stesso modo prevede che la relazione fra la variabile Y e la variabile X_2 (capitale culturale) dipende dalla variabile Z. In termini sostantivi, se γ_{11} è positivo, l'effetto del genere sul risultato al compito di comprensione del testo (l'effetto di X_1 su Y) è maggiore in classi con insegnanti con un'alta esperienza, e viceversa; se γ_{21} è positivo, l'effetto del capitale culturale sul risultato al compito di comprensione del testo (l'effetto di X_2 su Y) è maggiore in classi con insegnanti con molta esperienza, se invece è negativo il capitale culturale avrà un effetto relativamente minore sul risultato al test.

I termini di errore u_{0j} , u_{1j} e u_{2j} , sono i residui (*random residual errors*) al livello delle classi, ossia i residui o errori delle equazioni di secondo livello. Si assume che i termini di errore abbiano media zero e siano indipendenti dagli errori di primo livello. La varianza dei residui u_{0j} , u_{1j} e u_{2j} è specificata rispettivamente come σ^2_{u0} , σ^2_{u1} e σ^2_{u2} ; la covarianza tra gli errori è assunta essere diversa da zero e denominata σ_{u01} , σ_{u02} e σ_{u12} .

Le equazioni 1, 2, 3 e 4 che compongono il nostro modello possono essere riscritte come singola equazione sostituendo le equazioni 2, 3 e 4 in 1. Otteniamo:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + \gamma_{01}Z_j + \gamma_{11}Z_jX_{1ij} + \gamma_{21}Z_jX_{2ij} + u_{1j}X_{1ij} + u_{2j}X_{2ij} + u_{0j} + e_{ij} \quad (5)$$

L'equazione 5, a differenza di tutte le precedenti, contiene dei termini di interazione fra le variabili di primo livello e l'unica variabile di secondo livello inserita nel nostro modello. Tali interazioni, formalizzate con $\gamma_{11}X_{1ij}Z_j$ e $\gamma_{21}X_{2ij}Z_j$, sono implicite nel processo di sostituzione delle equazioni 2, 3 e 4 in 1 ed esprimono la relazione fra X e Y come funzione di Z nelle equazioni di secondo livello nel tentativo di spiegare la variazione dei coefficienti di regressione casuali (β_j). In altre parole, la variabile Z agisce da moderatore della relazione. Tale effetto è il cosiddetto *cross-level interaction term*. Da notare che anche i residui u_{1j} e u_{2j} sono connessi con X_{ij} . Tale situazione è descritta come eteroschedasticità nel modello di regressione classico in quanto gli errori, essendo moltiplicati per le variabili esplicative, non sono indipendenti da queste ultime e saranno diversi a seconda dei differenti valori che assume X. Nuovamente, non viene rispettato uno degli assunti fondamentali della regressione classica – l'omoschedasticità – sottolineando ancora una volta come quest'ultima tecnica sia inadatta a trattare dati annidati gerarchicamente.

L'equazione 5 è composta da elementi di due tipi: una parte fissa ed una casuale (*random*). Gli elementi " $\gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + \gamma_{01}Z_j + \gamma_{11}X_{1ij}Z_j + \gamma_{21}X_{2ij}Z_j$ " rappresentano la parte fissa, ossia i coefficienti γ che rimangono costanti per tutte le unità della variabile di secondo livello. Mentre, la parte *random* è rappresentata dagli elementi " $u_{1j}X_{1ij} + u_{2j}X_{2ij} + u_{0j} + e_{ij}$ " e contiene i termini di errore *random* del modello, ossia la componente variabile dell'intera equazione.

Generalizzando, gli effetti stimati dal modello di regressione multilivello sono:

- parte fissa
 - γ_{00} : è l'intercetta, ossia il valore medio di Y (tra tutti i gruppi e gli individui) quando X e Z sono pari a zero;
 - γ_{10} : è l'effetto del predittore a livello individuale X quando Z assume valore zero;
 - γ_{01} : è l'effetto del predittore di secondo livello Z;
 - γ_{11} : è l'effetto di interazione fra il predittore a livello individuale e quello a livello dei gruppi, cioè l'effetto di Z sull'effetto di X su Y;
- parte casuale
 - σ^2_e : è la varianza del livello d'errore inferiore e_{ij} , ossia la varianza intra-classe controllando per l'effetto di X;
 - σ^2_{u0} : è la varianza condizionata dell'intercetta di primo livello nelle unità di secondo livello, ossia la varianza condizionata dell'intercetta controllando per X e Z (esprime la variabilità tra i gruppi di livello-2 per la parte relativa alla sola intercetta);
 - σ^2_{u1} : è la varianza condizionata del coefficiente di regressione di primo livello nelle unità di secondo livello (esprime la variabilità tra i gruppi di livello-2 dell'effetto di X su Y controllando per Z);
 - σ_{u01} : è la covarianza condizionata fra intercetta e coefficiente di regressione di primo livello.

Se nel modello sono presenti almeno due variabili esplicative *random*, alle quali è permesso variare fra le unità di secondo livello, avremo anche il termine:

- σ_{u12} : la covarianza condizionata dei coefficienti di regressione di primo livello.

2.2.3 Il modello nullo

Il ricorso all'analisi multilivello ha senso solo in presenza di dati annidati (*nested*), ossia nei quali è individuabile una struttura gerarchica. Come detto in precedenza, ciò produce un fattore di disturbo riguardante la dipendenza delle osservazioni: queste non possono essere più assunte come indipendenti. Il grado di similarità fra le diverse osservazioni può essere calcolato attraverso il coefficiente di correlazione intraclasse ($\rho = \rho$), che fornisce una misura del grado di omogeneità tra osservazioni appartenenti allo stesso gruppo (annidate all'interno di una unità macro). Il coefficiente di correlazione intraclasse è il rapporto fra la varianza attribuibile ai gruppi (*between*) e la varianza totale. Al fine di calcolare il coefficiente di correlazione intraclasse è utile ricorrere al cosiddetto "modello nullo" (*intercept-only model*). Tale modello non contiene nessuna variabile esplicativa (nessuna X) ma soltanto l'intercetta e i rispettivi termini di errore. Non avendo predittori, l'equazione assume la forma:

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \quad (6)$$

Il modello nullo nell'equazione 6 non spiega nessuna varianza in Y, semplicemente decompone la varianza totale presente nella Y in due componenti indipendenti: la varianza di primo livello (σ^2_e) dei residui di livello-1 e la varianza di secondo livello σ^2_{u0} dei residui di livello-2. In altre parole, il modello nullo ci restituisce una stima della variabilità della variabile oggetto di studio all'interno delle unità di livello superiore. In genere, prima di impostare un'analisi multilivello, che può richiedere il dispendio di ingenti risorse cognitive e di tempo, è preferibile specificare il modello nullo in modo da calcolare la percentuale di varianza dovuta alle unità di secondo livello sulla cui base valutare se proseguire o meno.

Con i parametri stimati attraverso il modello nullo è possibile calcolare il coefficiente di correlazione intraclasse (ρ) che indica la proporzione di varianza spiegata (riprodotta) dalle unità di secondo livello la cui formula è:

$$\rho = \sigma^2_{u0} / \sigma^2_{u0} + \sigma^2_e$$

2.2.4 Il modello ad intercetta casuale e con coefficiente di regressione casuale

Abbiamo visto come attraverso l'analisi multilivello sia possibile introdurre la variabilità nei coefficienti di regressione permettendo ad essi di variare all'interno dei gruppi che costituiscono le unità di secondo livello. L'obiettivo dell'analisi multilivello, ovviamente, è spiegare tale variabilità inserendo nel modello variabili esplicative di secondo livello⁶. Ciò è abbastanza intuitivo tenendo presente che se i coefficienti mostrano variabilità a livello di gruppo (livello-2) vuol dire che vi è qualche proprietà contestuale o strutturale inerente ai gruppi capace di dar conto di tale variazione. Per esempio, consideriamo tre classi scolastiche diverse in cui somministriamo un test volto alla valutazione della conoscenza della matematica inserendo come variabile esplicativa il quoziente intellettivo. Se nelle tre equazioni di regressione che ne conseguono otteniamo tre valori diversi delle intercette – poniamo che il valore riferito alla terza classe è maggiore degli altri due – vuol dire che nella scuola 3, a parità di QI, gli studenti mostrano una conoscenza maggiore della matematica. In definitiva, la variabilità delle intercette non è altro che l'effetto di gruppo, ossia la scuola 3 avrà qualche caratteristica che fa sì che, a parità di QI, gli studenti abbiano una conoscenza superiore della matematica⁷.

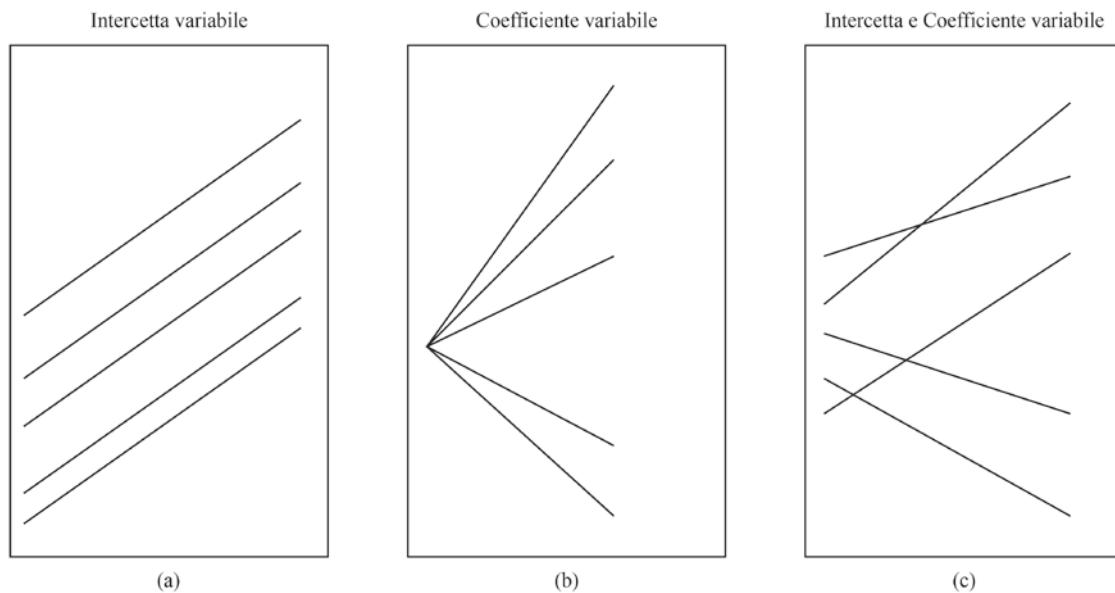
Nei modelli multilivello vi sono essenzialmente due fonti di variabilità – l'intercetta e il coefficiente di regressione – che danno origine a tre modelli diversi: ad intercetta casuale, con coefficiente di regressione casuale ed intercetta costante, con intercetta e coefficienti di regressione casuali (fig.2.4). Il modo più semplice per introdurre la variabilità dei parametri è permettere all'intercetta di variare all'interno delle unità di secondo livello (fig.2.4a). Per un modello ad intercetta variabile senza variabili esplicative di secondo livello e con un solo predittore a livello individuale l'equazione è:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + u_{0j} + e_{ij} \quad (7)$$

⁶ Anche i predittori di primo livello sono in grado di dar conto della variabilità di secondo livello se la variabile di primo livello inserita nel modello ha una diversa distribuzione all'interno dei gruppi che costituiscono il livello-2. Ciò non succede per i predittori di secondo livello che in nessun modo possono contribuire alla riduzione della varianza di primo livello.

⁷ Sta al ricercatore poi individuare quali fra le infinite proprietà permettono di dar conto di tale relazione.

Fig.2.4 – modello ad intercetta casuale, intercetta fissa e coefficiente di regressione casuale, ad intercetta e coefficienti di regressione casua



Gli assunti fondamentali del modello sono: i termini di errore (u_{0j} , e_{ij}) sono indipendenti fra di loro ed hanno media zero, si assume che derivino da una popolazione normalmente distribuita e che la varianza dei residui di primo livello (e_{ij}) è costante fra le unità di secondo livello.

Il termine u_{0j} indica di quanto le diverse intercette delle unità di secondo livello si discostano dall'intercetta generale, può essere quindi interpretato come residuo di secondo livello dopo aver controllato per la variabile esplicativa X o piuttosto come l'effetto delle unità di secondo livello dopo aver controllato per X. Dal momento che u_{0j} contiene la parte di variabilità della variabile dipendente che non è stata spiegata dal regressore X, tale modello contiene una parte di varianza di secondo livello non spiegata. La suddivisione della varianza non spiegata fra i diversi livelli di analisi (non per forza due) è l'essenza dell'analisi multilivello (Snijder e Bosker 1999). A questo punto il ricercatore può pensare di inserire nel modello variabili esplicative di livello superiore Z per cercare di ridurre tale varianza residua⁸. In tal caso, per un modello ad intercetta variabile con una variabile esplicativa di secondo livello ed un predittore a livello individuale l'equazione risulta essere:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{01}Z_j + u_{0j} + e_{ij} \quad (8)$$

Il secondo modello che prendiamo in esame (fig.2.4b) prevede invece che a variare sia soltanto il coefficiente di regressione mentre l'intercetta rimane costante fra le unità di secondo livello. È un modello abbastanza inusuale, di cui non vi è quasi traccia in letteratura⁹, in quanto postula che non vi sia nessuna differenza media nella variabile indipendente fra le unità di secondo livello (l'intercetta è costante) ma che la relazione fra X e Y sia più intensa o meno intensa a seconda dei gruppi di livello-2.

⁸ Come detto, non solo le variabili di secondo livello ma anche quelle di primo livello possono contribuire alla riduzione della varianza di secondo livello.

⁹ In base alle nostre conoscenze, vi è un solo riferimento in Gelman e Hill (2006).

Per esempio, in una classe scolastica l'effetto del capitale culturale posseduto dall'alunno influisce maggiormente sui risultati scolastici ottenuti rispetto ad un'altra classe dove tale variabile ha un effetto relativamente inferiore. Tale modello dà origine ad un diagramma a dispersione con rette di regressione "a megafono" (fig.2.4b) Pur trattandosi di un modello poco diffuso, per completezza di esposizione se ne fornirà l'equazione:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + u_{1j}X_{1ij} + e_{ij} \quad (9)$$

Il termine u_{1j} contiene i residui dei coefficienti di regressione delle unità di secondo livello dal coefficiente di regressione medio. Nell'equazione tale parametro è moltiplicato per i valori di X_{ij} generando eteroschedasticità poiché i valori dei residui non sono indipendenti dalla variabile esplicativa X . Il termine $u_{1j}X_{1ij}$, in questo caso rappresenta l'interazione fra l'unità di secondo livello e la variabile indipendente X .

Più consueto è il modello con intercetta e coefficiente angolare casuali (fig.2.4c). In questo caso le unità di secondo livello differiscono sia in riferimento al valore medio della variabile indipendente (intercetta casuale) sia riguardo al coefficiente di regressione determinando una situazione in cui nei gruppi di secondo livello la variabile esplicativa X può avere un effetto su Y di intensità variabile a seconda del gruppo di riferimento. Ritornando al primo esempio, i livelli medi di conoscenza delle matematica non solo variano a seconda della classe scolastica presa in considerazione ma il QI esercita effetti di maggiore o minore intensità sulla variabile Y a seconda della classe presa in considerazione. In altre parole, alle intercette e ai coefficienti di regressione è permesso variare fra le unità di secondo livello. Rispetto al primo modello, quello con la sola intercetta casuale, qui abbiamo un ulteriore coefficiente *random* associato al coefficiente di regressione. In tal caso, l'equazione diventa:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + u_{1j}X_{1ij} + u_{0j} + e_{ij} \quad (10)$$

Così come per l'analisi di regressione classica, l'obiettivo dell'approccio multilivello è spiegare la varianza della variabile dipendente. La differenza con l'analisi di regressione classica risiede nel fatto che la regressione multilivello produce una struttura della varianza più complessa. Dal momento che sia l'intercetta che i coefficienti di regressione sono casuali (*random*), si deve tentare di dar conto di tale variazione. In altre parole, una volta stabilito che i coefficienti variano a livello superiore bisogna tentare di spiegare tale variazione. Nel modello a sola intercetta casuale abbiamo visto come, inserendo un predittore di secondo livello, si può tentare di spiegare la varianza associata all'intercetta. Lo stesso, con qualche accorgimento, può essere fatto nel tentativo di dar conto della variabilità dei coefficienti di regressione. Presenteremo prima le equazioni di regressione per poi fornirne la spiegazione.

Per un modello ad intercetta e coefficiente di regressione casuali con una variabile esplicativa di primo livello ed una di secondo livello le equazioni sono:

primo livello

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \varepsilon_{ij} \quad (11)$$

secondo livello

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j} \quad (12)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_j + u_{1j} \quad (13)$$

Unendo le due equazioni otteniamo:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{01}Z_j + \gamma_{11}Z_j X_{1ij} + u_{1j}X_{1ij} + u_{0j} + e_{ij} \quad (14)$$

Notiamo come nell'equazione congiunta vi sia un parametro ($\gamma_{11} Z_j X_{1ij}$) apparentemente non presente nelle equazioni di primo e secondo livello. Ricordiamo che i coefficienti di primo livello β_{0j} e β_{1j} sono trattati come variabili dipendenti nelle equazioni di secondo livello – con una propria intercetta e un proprio coefficiente di regressione – al fine di predirne i valori attraverso la variabile di secondo livello Z . Quindi, tenendo presenti le equazioni in 11, 12 e 13 il modello potrebbe essere riscritto come:

$$Y_{ij} = (\gamma_{00} + \gamma_{01}Z_j + u_{0j}) + (\gamma_{10} + \gamma_{11}Z_j + u_{1j}) X_{1ij} + e_{ij} \quad (15)$$

Riscrivendo l'equazione in 14 in modo che le parti fisse dell'equazione vengano per prime, seguite dalle parti *random*, otteniamo l'equazione 15 e il termine di interazione $\gamma_{11} Z_j X_{1ij}$ che è il cosiddetto *cross-level interaction term*. È attraverso l'inserimento della variabile di secondo livello e il termine di interazione conseguente che si cerca di spiegare la varianza del coefficiente di regressione casuale. Ricordiamo, infatti, che l'equazione 13, ossia quella che cerca di predire il coefficiente di regressione casuale attraverso una variabile esplicativa di secondo livello, sostanzialmente sta a significare che la relazione fra la variabile X e Y dipende da Z .

2.2.5 La complessità delle equazioni di regressione multilivello

I modelli di regressione multilivello risultano essere notevolmente più complessi dei classici modelli di regressione mono-livello. Una serie di condizioni influiscono su tale complessità. Innanzitutto, il numero dei parametri è notevolmente superiore nei modelli multilivello. Inoltre, tali modelli, per le ragioni viste in precedenza, contengono effetti di interazione fra le variabili, ponendo in tal modo numerosi problemi di interpretazione. Infine, contengono numerosi livelli di varianza residua a fronte dell'unico termine di errore presente nella regressione classica (Hox 2010).

Illustreremo tale complessità servendoci di un esempio. Prendiamo in considerazione un modello dove oltre la variabile indipendente abbiamo cinque variabili esplicative. L'equazione di regressione mono-livello assumerà tale forma:

equazione mono-livello

$$Y_{ij} = \beta_0 + \beta_1X_{1i} + \beta_2X_{2i} + \beta_3X_{3i} + \beta_4X_{4i} + \beta_5X_{5i} + \epsilon_i \quad (16)$$

Ora, consideriamo lo stesso modello ma utilizzando un approccio multilivello. Conservando lo stesso numero di variabili, per un modello multilivello (con due livelli) con intercetta casuale, tre variabili indipendenti di primo livello con coefficienti casuali e due variabili indipendenti di secondo livello l'equazione diventa:

primo livello

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{1ij} + \beta_{2j}X_{2ij} + \beta_{3j}X_{3ij} + \varepsilon_{ij} \quad (17)$$

secondo livello

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_{1j} + \gamma_{02}Z_{2j} + u_{0j} \quad (18)$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}Z_{1j} + \gamma_{12}Z_{2j} + u_{1j} \quad (19)$$

$$\beta_{2j} = \gamma_{20} + \gamma_{21}Z_{1j} + \gamma_{22}Z_{2j} + u_{2j} \quad (20)$$

$$\beta_{3j} = \gamma_{30} + \gamma_{31}Z_{1j} + \gamma_{32}Z_{2j} + u_{3j} \quad (21)$$

L'equazione completa risulta essere:

$$Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + \gamma_{30}X_{3ij} + \gamma_{01}Z_{1j} + \gamma_{02}Z_{2j} + \gamma_{11}Z_{1j}X_{1ij} + \gamma_{12}Z_{2j}X_{1ij} + \gamma_{21}Z_{1j}X_{2ij} + \gamma_{22}Z_{2j}X_{2ij} + \gamma_{31}Z_{1j}X_{3ij} + \gamma_{32}Z_{2j}X_{3ij} + u_{0j} + u_{1j}X_{1ij} + u_{2j}X_{2ij} + u_{3j}X_{3ij} + \varepsilon_{ij} \quad (22)$$

Anche visivamente, l'equazione 22 è notevolmente più complessa della 16 nonostante il numero di variabili sia rimasto invariato. In un modello multilivello il numero dei parametri può facilmente diventare molto ampio creando sia problemi di convergenza del modello che problemi di interpretazione divenendo cognitivamente ingestibile. In particolare, la complessità è dovuta principalmente all'introduzione dei coefficienti di regressione casuali associati all'introduzione di variabili di secondo livello. Infatti, per ogni coefficiente *random* cui si cerca di dar conto attraverso variabili di secondo livello si aggiunge, oltre al termine di errore associato ad ogni coefficiente di regressione, il termine di interazione fra tale coefficiente e le variabili esplicative di secondo livello. In altre parole, se abbiamo due coefficienti di regressione *random* cui si cerca di dar conto attraverso due variabili esplicative di secondo livello, oltre a questi quattro parametri e i termini di errore, si aggiungeranno ben quattro termini di interazione fra tali variabili, come spiegato nel paragrafo precedente. In particolare, l'equazione 22 contiene in tutto 23 parametri (a fronte dei sette dell'equazione mono-livello 25) ossia: un'intercetta, un residuo di primo livello, tre coefficienti di regressione di primo livello, la varianza residua a livello delle unità di secondo livello, tre varianze residue dei coefficienti di regressione di secondo livello, tre covarianze fra l'intercetta con i coefficienti di regressione, tre covarianze fra i coefficienti di regressione, due coefficienti di regressione di secondo livello, infine sei termini di interazione *cross-level*. In generale, date p variabili esplicative di primo livello e q variabili esplicative di secondo livello, il numero di parametri stimato è dato dalla lista presente in tab.2.4 (Hox 2010). Date tutte queste complicazioni, numerosi manuali (Hox 2010; Snijders e Bosker 1999) consigliano di partire sempre dal modello più semplice e di sviluppare il modello nella maniera meno complessa possibile evitando di aggiungere inutili complicazioni attraverso l'aggiunta non teoricamente giustificata di coefficienti casuali.

In assenza di teorie, una procedura esplorativa *bottom-up* volta a valutare la significatività degli effetti casuali del modello potrebbe aiutare. In questo caso, si parte dal modello nullo per valutare la proporzione di varianza *between* sulla varianza totale. Se la porzione di varianza riprodotta dalle unità di secondo livello è soddisfacente (il limite minimo stabilito in letteratura è 5%, $p = 0,05$) si aggiungono i predittori di primo livello permettendo alle intercette di variare all'interno delle unità di secondo livello. Dopo aver valutato la significatività di ogni predittore e i cambiamenti nella varianza di primo e secondo livello, si aggiungono i predittori di secondo livello valutando se contribuiscono a ridurre la varianza *between* della variabile dipendente. Successivamente, si permette ai coefficienti di

regressione di variare nelle unità di secondo livello valutando la significatività di tale variazione. Infine, si cerca di dar conto della variazione dei coefficienti di regressione nelle unità di secondo livello introducendo i termini di interazione *cross-level*.

Tab.2.4 – Numero parametri stimati in un modello multilivello

Parametri	Numero
Intercetta	1
Varianza dei residui livello-1	1
Coefficienti di regressione fissi per i predittori livello-1	p
Varianza residui livello-2	1
Varianze residue livello-2 dei coefficienti di regressione	p
Covarianze livello-2 fra l'intercetta con i coefficienti di regressione	p
Covarianze livello-2 fra i coefficienti di regressione	$p(p - 1)/2$
Coefficienti di regressione fissi per i predittori livello-2	q
Coefficienti di regressione fissi per le interazioni <i>cross-level</i>	$p \times q$

2.3 I modelli ad equazioni strutturali

Fra la fine degli anni '60 e l'inizio degli anni '70 Jöreskog, uno statistico psicometrico di origini svedesi, propone un nuovo approccio all'analisi dei dati. Riunendo in una sola formulazione tre differenti discipline delle scienze sociali, ossia la sociologia, la psicometria e la psicologia, produce una feconda sintesi fra l'analisi fattoriale, la *path analysis* e i modelli di equazioni simultanee. Il vero punto di forza dell'approccio – che ha permesso di risolvere due dei principali problemi che affliggevano i tradizionali modelli di *path analysis* e i modelli ricorsivi degli econometrici – risiede nella possibilità di includere costrutti latenti (Lazarsfeld 1954) e di tener conto dell'errore di misurazione delle variabili. I modelli ad equazioni strutturali intendevano fornire una risposta a due dei più importanti interrogativi presenti nelle scienze sociali. Il primo riguarda il problema della “misurazione” riferito sia al fatto che alcune variabili non possono essere misurate direttamente in quanto rappresentano concetti teorici non direttamente osservabili, sia in riferimento ai concetti di validità ed attendibilità delle misure, ovvero l'imprecisione degli strumenti volti a rilevare le variabili. Il secondo interrogativo riguarda il problema della causalità e la consapevolezza che ogni teoria scientifica si basa sull'elaborazione di nessi causali. Tale approccio permette in ultima analisi di congiungere i due interrogativi consentendo di instaurare “nessi causali” fra variabili latenti¹⁰ (Bollen 1989). Fin dall'inizio, la psicometria, attraverso gli studi di Spearman (1904) sul quoziente intellettivo, si è posta il problema della presenza delle variabili latenti individuando nell'analisi fattoriale la tecnica privilegiata per l'individuazione di fattori latenti volti a spiegare la struttura delle correlazioni di un certo numero di variabili osservate. Parallelamente, tale disciplina, in riferimento al problema della misurazione, portava avanti la ricerca sui concetti di validità e attendibilità. Contemporaneamente, le discipline economiche si ponevano il problema delle relazioni causali attraverso la messa a punto dei modelli di equazioni simultanee. Ma da tali modelli rimanevano fuori i concetti di variabile latente e di errore di misurazione, in quanto ritenuti irrilevanti per l'econometria, disciplina che ricorreva all'utilizzo di dati economici considerati immuni dagli errori di misurazione (Goldberger 1972). A metà degli anni sessanta, Duncan (1966), riprendendo i lavori degli anni trenta del genetista Wright, rimasti sconosciuti nelle scienze sociali, diffuse i modelli di *path analysis* fra i sociologi. Fu all'inizio degli anni settanta che tali tradizioni di ricerca subirono un processo di avvicinamento per merito soprattutto di Jöreskog per la psicometria, Goldberger per l'econometria e Duncan per la sociologia. L'apice fu il seminario organizzato dallo stesso Goldberger nel 1970 nell'Università del Wisconsin, dove Jöreskog presentò il programma Lisrel che permise di integrare in un unico *software* le problematiche relative all'analisi causale, all'analisi fattoriale e ai modelli di misurazione.

Fondamentale nei modelli ad equazioni strutturali è la distinzione fra il modello di misura e il modello strutturale. Attraverso il modello di misura si stabilisce mediante quali variabili manifeste si stimano le variabili latenti valutandone attendibilità e validità. Il modello strutturale, all'inverso, specifica le relazioni fra le variabili latenti stimandone l'entità e la quota di varianza residua.

A titolo di esempio, pensiamo di dover stimare un modello strutturale valutandone gli “effetti causali” composto da due latenti esogene (ξ_1, ξ_2) e due endogene (η_1, η_2). Essendo le quattro variabili di tipo latente, si dovrà fare ricorso ad un certo numero di variabili manifeste al fine di stimarne i valori. Poniamo che per ogni latente esogena facciamo ricorso a tre variabili esogene manifeste (x) e per ogni

¹⁰ C'è bisogno di una duplice precisazione terminologica. Siamo perfettamente consapevoli che in assenza di disegni di ricerca sperimentali non si può in nessun modo parlare di “relazioni causali” (Kline, 2011) e che “variabile latente” costituisce un ossimoro essendo la variabile sempre frutto di una definizione operativa con l'obiettivo di renderla rilevabile.

latente endogena utilizziamo tre endogene manifeste (y). Il modello può essere rappresentato attraverso le equazioni strutturali.

Le equazioni del modello strutturale possono scriversi¹¹:

$$\eta_1 = \gamma_{11} \xi_1 + \zeta_1$$

$$\eta_2 = \beta_{21} \eta_1 + \gamma_{22} \xi_2 + \zeta_2$$

Le equazioni del modello di misura per le latenti esogene sono:

$$x_1 = \lambda_{11}^x \xi_1 + \delta_1$$

$$x_2 = \lambda_{21}^x \xi_1 + \delta_2$$

$$x_3 = \lambda_{31}^x \xi_1 + \delta_3$$

$$x_4 = \lambda_{42}^x \xi_2 + \delta_4$$

$$x_5 = \lambda_{52}^x \xi_2 + \delta_5$$

$$x_6 = \lambda_{62}^x \xi_2 + \delta_6$$

Le equazioni del modello di misura per le latenti endogene risultano essere:

$$y_1 = \lambda_{11}^y \eta_1 + \varepsilon_1$$

$$y_2 = \lambda_{21}^y \eta_1 + \varepsilon_2$$

$$y_3 = \lambda_{31}^y \eta_1 + \varepsilon_3$$

$$y_4 = \lambda_{42}^y \eta_2 + \varepsilon_4$$

$$y_5 = \lambda_{52}^y \eta_2 + \varepsilon_5$$

$$y_6 = \lambda_{62}^y \eta_2 + \varepsilon_6$$

La prima equazione del modello strutturale esprime che la variabile latente endogena è funzione di una variabile latente esogena (ξ_1) più un errore stocastico (ζ_1); allo stesso modo nella seconda equazione η_2 è dipendente dalla prima variabile latente endogena e dalla seconda variabile latente esogena (ξ_2) più un errore stocastico (ζ_2). Le equazioni del modello di misurazione per le variabili latente endogene ed esogene differiscono solo in riferimento ai simboli utilizzati. I coefficienti lambda, che esprimono il peso fattoriale della latente sulla manifesta, differiscono soltanto per l'apice che assumono indicando se appartengono a manifeste esogene (apice X) o manifeste endogene (apice Y). Per il resto descrivono che ciascuna variabile manifesta è funzione della specifica variabile latente (esogena ξ o endogena η) e del rispettivo termine di errore (δ per le esogene ed ε per le endogene).

2.3.1 Il modello strutturale e il modello di misurazione

Abbiamo detto che la particolarità dei modelli ad equazioni strutturali risiede nella possibilità di unire all'interno di un unico modello diverse tradizioni di ricerca consentendo di specificare contemporaneamente un modello di misurazione, assimilabile all'analisi fattoriale, ed un modello strutturale, riferibile alla *path analysis*. Ma, a differenza dei classici modelli strutturali e di *path analysis*

¹¹ Adesso ed in seguito si assumerà sempre che le variabili siano espresse in scarti dalla loro media, in tal modo viene a scomparire l'intercetta, semplificando le formulazioni algebriche.

diffusi fra gli econometristi e i sociologi che consentivano il ricorso alle sole variabili manifeste, il modello di Jöreskog permette di integrare costrutti latenti, i quali, spesso, sono quelli che hanno una effettiva rilevanza teorica.

Un modello strutturale definisce i legami strutturali che collegano le diverse variabili quantificandone sia gli "effetti causali" che l'ammontare di varianza residua delle variabili endogene, ossia quella non spiegata dal modello (Jöreskog e Sörbom 1979).

Al fine di presentare il modello strutturale generale (Bollen 1989), ritorniamo alle equazioni del modello strutturale presentate in precedenza. Queste possono essere riscritte rendendo espliciti i parametri fissati a zero, ossia quei parametri riferiti all'assenza di relazioni. Le equazioni del modello strutturale diventano:

$$\begin{aligned}\eta_1 &= 0\eta_1 + 0\eta_2 + \gamma_{11}\xi_1 + 0\xi_2 + \zeta_1 \\ \eta_2 &= \beta_{21}\eta_1 + 0\eta_2 + 0\xi_1 + \gamma_{22}\xi_2 + \zeta_2\end{aligned}$$

L'equazione generale del modello strutturale in cui le variabili endogene siano pari ad m e le variabili esogene paria a n può essere scritta come:

$$\eta = \mathbf{B}\eta + \mathbf{\Gamma}\xi + \zeta$$

dove abbiamo:

- i tre vettori delle variabili endogene, esogene e degli errori con η che è il vettore delle m variabili latenti endogene, ξ che il è il vettore delle n variabili esogene e ζ , ossia il vettore dei residui o errori delle variabili endogene;
- le due matrici dei coefficienti strutturali con $\mathbf{B}$ (beta) che rappresenta la matrice dei coefficienti strutturali fra le variabili latenti endogene e $\mathbf{\Gamma}$ (gamma) la matrice dei coefficienti strutturali fra le variabili latenti esogene ed endogene. $\mathbf{B}$ è una matrice di ordine $m*m$, ossia una matrice quadrata di dimensioni pari al numero delle variabili endogene, mentre $\mathbf{\Gamma}$ è di ordine $m*n$.

Oltre a ciò, bisogna ulteriormente considerare la presenza di altre due matrici, ossia la matrice di varianze e covarianze fra le variabili latenti esogene e la matrice di varianze e covarianza fra i residui delle equazioni (ossia il termine di errore ζ). La prima si indica con Φ (phi) ed è di ordine $n*n$, dove n è sempre il numero delle variabili esogene, mentre la seconda con Ψ (psi) di ordine $m*m$, dove m rappresenta il numero delle variabili endogene. Essendo matrici di varianze e covarianze sono entrambe matrici quadrate e simmetriche (Bollen 1989; Jöreskog e Sörbom 1979; Kline 2011).

La stessa procedura può essere applicata per il modello di misurazione delle variabili endogene ed esogene. Partiamo dal primo.

Le equazioni del modello di misurazione per le variabili endogene presentate nel paragrafo precedente possono essere riscritte rendendo espliciti i parametri fissati a zero. Le equazioni del modello di misurazione per variabili endogene così diventano:

$$\begin{aligned}y_1 &= \lambda_{11}^y \eta_1 + 0\eta_2 + \varepsilon_1 \\ y_2 &= \lambda_{21}^y \eta_1 + 0\eta_2 + \varepsilon_2 \\ y_3 &= \lambda_{31}^y \eta_1 + 0\eta_2 + \varepsilon_3 \\ y_4 &= 0\eta_1 + \lambda_{42}^y \eta_2 + \varepsilon_4\end{aligned}$$

$$y_5 = 0\eta_1 + \lambda_{52}^y \eta_2 + \varepsilon_5$$

$$y_6 = 0\eta_1 + \lambda_{62}^y \eta_2 + \varepsilon_6$$

L'equazione per il modello generale di misurazione delle variabili endogene risulta essere:

$$y = \Lambda_y \eta + \varepsilon$$

In questa equazione sono presenti tre vettori ed una matrice, ovvero:

- i tre vettori delle variabili endogene osservate, endogene latenti e gli errori indicati rispettivamente con y , η e ε . I vettori y e ε contengono tanti elementi quante sono le variabili osservate; il vettore η è di ordine pari al numero di variabili endogene presenti nel modello;
- la matrice riferita ai coefficienti strutturali (in questo caso definibili come pesi fattoriali) riguardanti le relazioni fra le variabili osservate e le variabili latenti, ossia la matrice indicata con Λ_y (lambda y) di ordine pari al prodotto del numero di variabili osservate endogene e il numero di variabili latenti endogene.

A ciò deve aggiungersi la matrice di varianze e covarianze degli errori di misura (ε) del modello indicata col simbolo Θ_ε (theta epsilon). È una matrice quadrata simmetrica di ordine pari al prodotto del numero totale di variabili manifeste endogene moltiplicato per se stesso.

In riferimento al modello di misurazione per le variabili esogene, questo è del tutto simile al modello appena presentato con l'unica differenza che il modello ha per oggetto variabili osservate esogene (x) e variabili latenti esogene (ξ). Per tale motivo riporteremo soltanto l'equazione generale riferita al modello di misurazione per variabili esogene. In questo caso l'equazione diventa:

$$x = \Lambda_x \xi + \delta$$

Anche qui, esattamente come prima, abbiamo tre vettori e una matrice:

- i vettori si riferiscono alle variabili esogene manifeste (x), alle variabili esogene latenti (ξ) e, come sempre, ad un termine di errore riferito alle variabili esogene osservate, denominato questa volta δ . Ovviamente, i vettori x e δ sono pari al numero di variabili osservate e ξ è pari al numero di variabili latenti esogene specificate;
- la matrice, denominata in questo caso Λ_x (lambda x), si riferisce ai coefficienti strutturali (o pesi fattoriali) che collegano le variabili latenti esogene alle manifeste esogene. La matrice è di ordine pari al prodotto del numero delle variabili osservate per il numero di variabili latenti.

Anche in questo caso abbiamo un'ulteriore matrice quadrata e simmetrica, indicata con il simbolo Θ_δ (theta delta), riferita alle varianze e covarianze degli errori di misura (δ) delle variabili osservate esogene, di ordine pari al prodotto del numero totale di variabili manifeste esogene moltiplicato per se stesso.

2.3.2 Analisi fattoriale e analisi causale

Abbiamo detto in apertura che l'originalità dell'approccio dei modelli ad equazioni strutturali risiede nelle opportunità che offre di far fronte a due dei principali problemi con i quali lo scienziato sociale deve costantemente confrontarsi: quello della misurazione e quello della causalità.

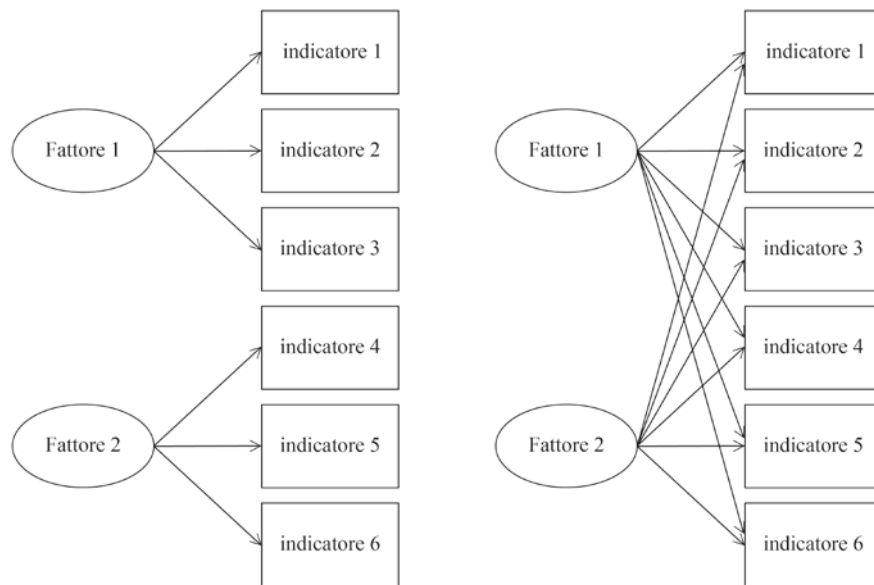
Al primo problema possono ricondursi essenzialmente due tematiche di ricerca entrambe di origine psicometrica – l'analisi fattoriale e i modelli di misurazione – mentre al secondo è ascrivibile la tradizione sociologica e la costruzione di modelli con la specificazione di percorsi causali che mettono in relazione una serie di variabili.

Di particolare rilevanza per le scienze sociali in generale e la sociologia in particolare sono le tecniche di analisi fattoriale e i modelli causali.

Obiettivo principale dell'analisi fattoriale è l'individuazione di dimensioni latenti, i fattori, sottostanti ad un insieme di variabili manifeste fra loro correlate ed assunte ad indicatori di tali dimensioni. Nella sua formulazione originale, l'obiettivo quindi è quello di "spiegare" le correlazioni fra le variabili attraverso il ricorso a fattori sottostanti dei quali le variabili rilevate sarebbero soltanto delle manifestazioni empiriche di qualcosa che le presuppone (Stevens 2012). La matrice di varianze e covarianze (o le correlazioni) fra le variabili osservate costituiscono i dati su cui si effettua tale analisi che si basa sul principio dell'indipendenza locale: individuando i fattori latenti e tenendoli costanti le correlazioni originali presenti in matrice fra le variabili osservate si annullano. Vi sono due possibili approcci all'analisi fattoriale. Nel primo caso, quello più comune, il ricercatore non ha sufficienti elementi su cui basarsi per effettuare la scelta relativa al numero di fattori latenti capaci di dar conto delle correlazioni fra le variabili rilevate né di individuare quali fra queste siano indicatori del primo fattore invece del secondo e così via. In questo caso l'analisi fattoriale è di tipo esplorativo ed è volta all'individuazione sia del numero di fattori sia delle relazioni che legano i fattori ai propri indicatori. Sono tali aspetti di indeterminatezza che hanno generato per decenni numerose critiche a tale tecnica sottolineandone gli aspetti eccessivamente soggettivi, ossia lasciati all'indiscrezione del ricercatore come i diversi modi di stimare le comunaltà fra le variabili e di estrarne i fattori così come le diverse tecniche di rotazione a disposizione (Stevens 2012). Diversa è la situazione in cui il ricercatore sulla base di precedenti teorie o indagini riesce a formulare delle ipotesi sul numero di fattori esistenti, sulle relazioni che sussistono fra tali fattori e le relazioni che li legano con gli indicatori rilevati. In questo caso, l'analisi fattoriale da esplorativa diventa confermativa. Oltre alla differenza nel tipo di approccio, esplorativo il primo, teoricamente orientato il secondo, anche la specificazione dei due modelli è nettamente diversa. In fig.2.5 sono presentati due modelli di analisi fattoriale. La differenza risiede nel fatto che il primo è di tipo confermativo e il secondo esplorativo. In riferimento a quest'ultimo, tale modello ha una struttura libera nel senso che è aperto ad includere tutte le possibili relazioni fra i fattori e tutte le possibili influenze dei fattori sulle variabili osservate. In teoria, ricorrendo ad un approccio esplorativo inizialmente non saremmo in grado nemmeno di fornire una rappresentazione grafica del modello in quanto il numero di fattori presenti è un'informazione di cui non disponiamo (Bollen 1989).

Il modello confermativo, al contrario, include una serie di ipotesi formulate dal ricercatore che assumono la forma di vincoli posti al modello dal ricercatore. Nel caso specifico, si è imposto che i fattori fossero in totale due e non correlati fra di loro e inoltre che alcune variabili x fossero influenzate da uno specifico fattore ma non da un altro.

Fig.2.5 – differenza analisi fattoriale confermativa ed esplorativa



Nel contesto dei modelli ad equazioni strutturali l'analisi fattoriale svolge un ruolo essenzialmente confermativo¹² con il ricercatore che formula preventivamente delle ipotesi per poi testarne la validità. Come visto in precedenza, il modello fattoriale è espresso dalla seguente equazione matriciale:

$$x = \Lambda_x \xi + \delta$$

L'ipotesi nel modello fattoriale è che la matrice empirica delle varianze e covarianze fra le variabili rilevate sia funzione di un certo numero di parametri. Tale matrice può essere espressa in funzione dei parametri del modello dalla seguente equazione:

$$\Sigma_{xx} = \Lambda_x \Phi \Lambda_x' + \Theta_\delta$$

L'equazione esprime che la varianza di ciascuna variabile osservata è scomponibile in due componenti fra loro ortogonali (indipendenti). La prima è detta comunaltà e rappresenta la porzione di varianza spiegata dai fattori comuni sottostanti; la seconda è detta unicità e rappresenta la parte residua non spiegata dai fattori comuni.

Come detto, nell'analisi fattoriale confermativa i dati di partenza sono costituiti dalla matrice di varianze e covarianze. Per tale ragione, un'applicazione rigorosa richiederebbe il ricorso soltanto a variabili cardinali (continue, quantitative). Tuttavia, sono state messe a punto procedure che permettono il ricorso anche a variabili di tipo ordinale e dicotomiche (Muthén 1984). In pratica, una specifica variabile ordinale manifesta può essere pensata come una misura grezza di una sottostante variabile continua latente. Se in ipotesi abbiamo due variabili osservate ordinali, poniamo X_1 e X_2 , e

¹² Essenzialmente confermativo e non esclusivamente in quanto gli aspetti esplorativi permangono in una certa misura. Per esempio, tutte le operazioni di miglioramento di un modello esulano da un contesto strettamente confermativo per assumere i contorni di un'analisi esplorativa.

abbiamo due variabili continue sottostanti (latenti) X_1^* e X_2^* che hanno una distribuzione bivariata normale, allora dai valori delle variabili rilevate possiamo stimare il coefficiente di correlazione che dovrebbe esistere fra le variabili latenti per poter avere quella distribuzione congiunta delle variabili ordinali osservate. In termini diversi, possiamo stimare il coefficiente di correlazione policorico (Pearson 1900) fra le continue latenti X_1^* e X_2^* , il quale non è semplicemente un coefficiente di correlazione fra variabili ordinali ma può essere considerato come la correlazione fra due variabili continue, in questo caso latenti. Se una delle due variabili X_1 e X_2 è continua il coefficiente assume il nome di coefficiente di correlazione poliseriale; nel caso di due variabili dicotomiche il coefficiente di correlazione è tetracorico. In questi casi, per evitare stime distorte degli errori standard e per il test di adattamento T del chi-quadrato il metodo di stima della massima verosimiglianza non può essere più utilizzato ma si deve ricorrere al metodo dei minimi quadrati ponderati (*weighted least squares*) (Jöreskog e Sörbom 1988).

In riferimento all'analisi causale i modelli ad equazioni strutturali costituiscono un notevole passo avanti rispetto alla tradizionale *path analysis* (Duncan 1966; Blalock 1967). Innanzitutto, i modelli ad equazioni strutturali permettono il ricorso a variabili latenti che spesso costituiscono costrutti teorici rilevanti ai fini dello sviluppo di teorie. Inoltre, la *path analysis* rappresenta una tecnica piuttosto che un modello, in quanto produce modelli non falsificabili. Ciò perché ogni equazione viene stimata separatamente e non esiste una stima generale del modello nel suo complesso. Infine, non permette la possibilità di prevedere che gli errori delle equazioni siano correlati. Nei modelli ad equazioni strutturali tale possibilità sussiste permettendo di quantificare la correlazione fra gli errori e valutando in tal modo la presenza di fattori esterni non presi in considerazione nel modello o di tenere sotto controllo l'effetto strumento¹³. L'obiettivo principale dell'analisi causale è la scomposizione dell'effetto generale esercitato da una variabile su un'altra in effetto totale ("causale") e in effetto spurio. In aggiunta, l'effetto totale può essere ulteriormente suddiviso in effetto diretto e effetto indiretto: il primo costituisce il legame diretto che lega due variabili, il secondo è invece la somma di tutti i percorsi indiretti che legano unidirezionalmente una variabile ad un'altra. Tale scomposizione è di fondamentale importanza in quanto ci permette, dopo aver suddiviso gli effetti in spurio e totale e quest'ultimo in diretto ed indiretto, di valutare in senso quantitativo e qualitativo il nesso strutturale che lega due variabili (Bollen 1989). Il modello causale generale, come già visto in precedenza, assume tale forma per le variabili latenti:

$$\eta = B\eta + \Gamma\xi + \zeta$$

Mentre, in presenza di sole variabili osservate, abbiamo:

$$Y = BY + \Gamma X + \zeta$$

Le due equazioni sono identiche, semplicemente nella seconda sostituiamo Y ad η e X a ξ in quanto vige la condizione $Y = \eta$ e $X = \xi$ ¹⁴.

Una distinzione fondamentale nei modelli causali è quella fra i cosiddetti modelli ricorsivi e non-ricorsivi (Bollen 1989). La differenza sussiste nel fatto che i modelli non-ricorsivi permettono di inserire nel modello effetti causali che retroagiscono dagli effetti alle cause, prevedendo in tal modo effetti

¹³ Variabili rilevate con lo stesso strumento di rilevazione potrebbero tendere a mostrare errori correlati.

¹⁴ Le variabili manifeste vengono considerate come latenti. In pratica vengono trattate come variabili con un solo indicatore.

reciproci fra le variabili. In termini più precisi, abbiamo un modello ricorsivo quando la matrice B (riferita ai parametri strutturali β) è una matrice subdiagonale (ossia con valori diversi da zero solo nel triangolo inferiore alla diagonale e con i restanti valori tutti pari a 0) e Ψ (ossia la matrice di varianze e covarianze fra gli errori) è una matrice diagonale (cioè con valori solo sulla diagonale e zero tutti i restanti). Nel caso della matrice B subdiagonale ciò sta a significare che esiste una sequenza gerarchica nelle variabili Y permettendo l'esistenza di relazioni solo fra variabili che precedono verso quelle che seguono ma non viceversa. Mentre, sostantivamente, dire che la matrice Ψ è diagonale significa che tale matrice contiene solo le varianze dei residui ma nessuna covarianza che equivale a dire che gli errori non sono correlati. Queste sono le due condizioni che definiscono un modello ricorsivo: errori non correlati e sequenza gerarchica delle Y . Nel caso in cui si permette ai residui di essere correlati abbiamo i cosiddetti modelli parzialmente ricorsivi, abbastanza frequenti nei modelli longitudinali (Bollen 1989; Kline 2011). Se invece la seconda condizione non è rispettata abbiamo i cosiddetti modelli non-ricorsivi, ovvero quei modelli che prevedono relazioni bi-direzionali fra le variabili sotto forma di *feed-back* (Y_1 influisce su Y_2 e viceversa) oppure la presenza di *loop* (Y_1 influisce su Y_2 che influisce su Y_3 che a sua volta influisce su Y_1)

2.3.3 Il modello completo

Il modello strutturale completo consiste nella fusione del modello di misura e del modello strutturale. Riportiamo per chiarezza le equazioni del modello generale composto, come detto, dall'unione delle tre equazioni precedentemente illustrate:

$$\eta = B\eta + \Gamma\xi + \zeta$$

$$y = \Lambda_y\eta + \varepsilon$$

$$x = \Lambda_x\xi + \delta$$

Quindi, un modello ad equazioni strutturali completo per essere interamente specificato necessita di otto matrici in totale: due riferite a coefficienti strutturali (B , Γ), due a pesi fattoriali (Λ_y , Λ_x) e quattro matrici di varianze covarianze (Φ , Ψ , Θ_ε e Θ_δ) di cui una riferita alle varianze e covarianze delle latenti esogene e tre riguardanti le varianze e covarianze dei residui.

Le equazioni poggiano sui seguenti assunti (Bollen 1989):

- $E(\eta) = E(\zeta) = E(\xi) = E(Y) = E(\varepsilon) = E(X) = E(\delta) = 0$, ovvero le variabili sono espresse come scarti dalle loro medie;
- $E(\xi \xi') = E(\eta \varepsilon') = E(\xi \delta') = E(\eta \delta') = E(\xi \varepsilon') = E(\zeta \varepsilon') = E(\zeta \delta') = E(\varepsilon \delta') = 0$, ovvero le variabili indipendenti e gli errori non sono correlati fra equazioni e all'interno della stessa equazione, così come gli errori delle diverse equazioni sono fra loro non correlati
- La matrice B deve essere positiva definita, ovvero le equazioni del modello che esprimono η devono essere indipendenti fra loro; in altri termini, nessuna endogena latente deve essere una combinazione lineare di altre variabili endogene.

2.3.4 Le fasi dei modelli ad equazioni strutturali

I dati empirici che costituiscono la base su cui modelli ad equazioni strutturali lavorano sono costituiti dalla matrice di varianze-covarianze, indicata con S , fra tutte le variabili rilevate. Su questa si stimano tutti i parametri del modello così come specificato dal ricercatore in base ai quali si produce una nuova matrice di varianze e covarianze, detta matrice teorica ed indicata con Σ . Successivamente, si confrontano le due matrici per valutare quanto la matrice teorica si distacchi da quella empirica: tanto più sono piccole le differenze fra le due matrici, tanto più il modello si adatterà ai dati. Sappiamo però che una specifica matrice osservata può essere compatibile con modelli causali fra loro differenti (il contrario non è vero nel senso che uno specifico modello può produrre una ed una sola matrice teorica). Quindi, in teoria i modelli ad equazioni strutturali non possono mai essere verificati definitivamente, in quanto esistono numerosi altri modelli teorici in grado di spiegare la struttura presente nella matrice di varianze e covarianze osservata, ma possono essere soltanto falsificati, ossia rifiutati nel caso di non adattamento ai dati. Se un modello si adatta ai dati, cioè la matrice teorica prodotta è ragionevolmente simile a quella osservata, non vuol dire che modello, così come specificato, sia l'unico esistente capace di dar conto della struttura dati. Potrebbero in teoria esistere modelli con un adattamento migliore, o in estrema ipotesi anche antagonisti a quello da noi specificato.

I modelli ad equazioni strutturali sono specificati secondo una procedura a cinque fasi (Bollen e Long 1993). La prima fase è quella della specificazione del modello teorico in cui il ricercatore, sulla base di conoscenze teoriche pregresse e di precedenti indagine empiriche, traduce in un sistema di equazioni strutturali la teoria da testare. In questa fase si definiscono quali sono le variabili osservate, quali le latenti che possono determinarne i valori e si stabiliscono i nessi strutturali fra le variabili. Come risultato finale si avranno una serie di parametri, fra tutti i possibili, che il ricercatore deciderà di stimare fissando a zero quelli non rilevanti in base alla costruzione teorica precedente.

Abbiamo poi la fase dell'identificazione del modello che riguarda se il modello è matematicamente stimabile dato in certo numero di parametri e di incognite: non possiamo stimare un modello che contiene più parametri degli elementi unici della matrice di covarianza (Bollen 1989).

Successivamente si procede alla stima effettiva dei parametri strutturali specificati. Il procedimento di stima è di carattere iterativo e si basa sul principio della massima verosimiglianza che ha lo scopo di minimizzare la distanza fra i dati teorici prodotti e quelli empirici a disposizione del ricercatore. Lo scopo è riprodurre la matrice di covarianze osservata nel modo più preciso possibile minimizzandone la distanza da quella teorica. Il processo iterativo si conclude quando ogni tentativo volto alla riduzione di tale distanza non produce nessun miglioramento nel modello in termini di riproduzione della matrice osservata.

Infine abbiamo la fase del controllo del modello, ossia del confronto fra la matrice attesa Σ prodotta attraverso i nuovi parametri stimati con quella empirica S . A questo punto sappiamo che la distanza fra le due matrici è la minima possibile dati il nostro modello e i dati osservati. Tuttavia, tale distanza fra le due matrici potrà comunque risultare eccessiva affinché si possa non rifiutare il modello da noi specificato. Se la differenza fra le due matrici sarà inferiore ad una certa soglia il modello non sarà rifiutato, al contrario il modello risulterà falsificato e si dovrà o modificarlo o respingerlo del tutto. Ultima fase è appunto quella della modifica o affinamento del modello. In alcuni casi, le modifiche da apportare saranno lievi, in altri saranno tali da risultare in modelli che poco somigliano a quelli pensati nella fase di specificazione. Anche nel caso in cui il modello si adattasse ai dati rimane sempre la

possibilità di apportare migliorie al modello pensato in origine; quindi, il ricercatore deve sempre chiedersi se il modello da lui pensato sia in qualche modo migliorabile.

2.4 Modelli multilivello ad equazioni strutturali

2.4.1 Perché applicare un approccio multilivello

I modelli ad equazioni strutturali costituiscono ormai un approccio di analisi ampiamente diffuso nell'ambito della sociologia e delle scienze sociali in generale (Ryu e West 2009; Rabe-Hesketh, Skrondal e Zheng 2007; Bollen 1989). Come visto in precedenza, tale approccio permette di dar conto di una struttura di varianze e covarianze di un insieme di variabili riproducendone le relazioni attraverso un numero inferiore di parametri rispetto a quelli originali (Bollen 1989; Kline 2011). Uno dei principali assunti dell'impianto teorico dei modelli ad equazioni strutturali (così come dai modelli di regressione multipla e, più in generale, da tutti i cosiddetti modelli lineari generalizzati) è l'indipendenza fra le osservazioni (Bentler e Chou 1987; Bollen 1989; Curran 2003). La violazione di tale assunto fondamentale conduce ad una serie di problematiche che inficiano gravemente l'analisi dei dati e i risultati teorici a cui si perviene (Dyer, Hanges e Hall 2005). Quando le osservazioni non sono indipendenti – e quindi siamo in presenza di una struttura gerarchica di dati (*nested data*) – l'analisi produce stime non efficienti. In particolare, gli errori standard risultano sistematicamente deflazionati inficiando in tal modo tutti i test di significatività (Moerbeek 2004; Pornprasertmanit, Lee e Preacher 2014; Raudenbush e Bryk 2002; Hox 2010; Snijders e Bosker 1999; Bauer 2003)¹⁵. Anche la varianza delle variabili risulta affetta da *bias*: la varianza delle variabili individuali in un modello che non tiene conto della struttura gerarchica dei dati risulta maggiore rispetto alla varianza delle stesse variabili in un modello multilivello. Ciò accade in quanto la varianza attribuibile al livello macro viene sommata a quella individuale. Allo stesso modo la varianza di variabili aggregate (composte dal valore medio derivato dai singoli valori individuali) è maggiore di quella delle stesse variabili in un modello multilivello in quanto una porzione della varianza delle variabili a livello individuale non viene stimata ma semplicemente sommata a quella di livello macro (Pornprasertmanit, Lee e Preacher 2014). Julian (2001), applicando un'analisi fattoriale confermativa a dati multilivello, dimostra gli effetti nefasti che derivano dall'ignorare la struttura gerarchica dei dati. In particolare, i parametri del modello risultano sovrastimati, gli errori standard sottostimati e la statistica del chi-quadrato sovrastimata portando in tal modo a rifiutare sistematicamente modelli che in realtà presentano un buon adattamento. Nonostante sia passato più di un decennio dallo studio di Julian, gli studi che applicano tecniche standard a dati multilivello sono ancora la maggioranza (vedi Pornprasertmanit, Lee e Preacher 2014 i quali presentano una corposa lista di studi che, pur basandosi su dati multilivello, non tengono conto di tale struttura). In termini sostantivi ciò si traduce nel fatto che un ricercatore può erroneamente rifiutare l'ipotesi nulla individuando relazioni fra variabili che in realtà non sussistono con ripercussioni sull'intero impianto teorico. Inoltre, ignorando la struttura multilivello dei dati si perde una notevole opportunità teorica. Non specificando nel modello relazioni fra variabili che appartengono a diversi livelli di analisi si perde l'opportunità di valutare gli effetti di gruppo o contestuali (Pornprasertmanit, Lee e Preacher 2014).

¹⁵ Un maggiore approfondimento riguardante le conseguenze negative che derivano dall'ignorare la struttura gerarchica presente nei dati è contenuto nel paragrafo dedicato all'analisi multilivello. Per ulteriori approfondimenti si può far riferimento a Julian 2001, Moerbeek 2004, Pornprasertmanit, Lee e Preacher 2014, Raudenbush e Bryk 2002, Hox 2010 e Snijders e Bosker 1999.

2.4.2 Pregi e limitazioni dei modelli di regressione multilivello e dei modelli ad equazioni strutturali

L'analisi multilivello, come visto in precedenza, è stata sviluppata, e si è diffusa, come tecnica adatta al trattamento di strutture di dati annidate gerarchicamente¹⁶. I modelli di regressione multilivello da questo punto di vista hanno sicuramente dei vantaggi rispetto ai modelli ad equazioni strutturali. In particolare, possono trattare strutture gerarchiche di dati e consentono di testare teorie che tengano conto degli effetti che i contesti, i gruppi e l'ambiente esercitano sull'individuo permettendo in tal modo di prendere in considerazione più livelli di analisi (Bauer 2003). I modelli multilivello permettono agli effetti dei predittori di variare all'interno delle unità di secondo livello – gli effetti non sono fissi ma *random* – (Hox 2010; Snijders e Bosker 1999). Sappiamo che i modelli ad equazioni strutturali nella loro formulazione classica possono incorporare soltanto un livello di analisi – in genere quello individuale.

Tuttavia, pur presentando i vantaggi visti sopra, i modelli di regressione multilivello scontano una serie di mancanze rispetto ai modelli ad equazioni strutturali. Innanzitutto, non è possibile valutare la bontà dell'intero modello attraverso il ricorso ad indici di adattamento così come avviene nei SEM (Bauer 2003). Non è possibile inglobare costrutti latenti attraverso il ricorso ad indicatori multipli (Curran 2003; Bauer 2003). Non è possibile specificare modelli di misurazione: nei modelli multilivello classici tutte le variabili sono “direttamente osservate” e quindi rilevate (assunte come) senza errore di misurazione¹⁷ (Hox 2013). Non tenendo conto dell'errore di misurazione si assume la perfetta validità di ogni singolo indicatore.

Quelle presentate costituiscono forti limitazioni se si considera che, da un lato, tener conto dell'errore di misurazione all'interno del modello ci permette di migliorare l'affidabilità dell'intero modello nonché l'attendibilità delle stime dei parametri del modello (Bollen 1989; Kline 2011), dall'altro, non bisogna dimenticare che molti processi sociali sono concepiti come processi strutturali operanti tra costrutti latenti (non osservati, non osservabili direttamente). Infine, i modelli multilivello non permettono di specificare modelli che incorporano articolate “catene causali” fra variabili, testando ad esempio complesse ipotesi di mediazione fra la variabile criterio e i suoi predittori (Bauer 2003; Preacher, Zhang e Zyphur 2011; Preacher, Zyphur e Zhang 2010). Le principali limitazioni dei modelli multilivello costituiscono all'opposto i principali punti di forza dei modelli ad equazioni strutturali come la possibilità di specificare modelli di misurazione incorporando nel modello variabili latenti, l'opportunità di testare relazioni causali complesse e, infine, la possibilità di testare il modello nella sua interezza (Bauer 2003).

2.4.3 La continuità fra i modelli di regressione multilivello e i modelli ad equazioni strutturali

Non sorprende che nel corso degli anni siano stati fatti numerosi tentativi nel cercare di integrare i due approcci, ossia quello dei modelli ad equazioni strutturali e i modelli multilivello (Goldstein e McDonald

¹⁶ Come visto, l'analisi multilivello, oltre a basi puramente statistiche, si fonda anche su esigenze di tipo teorico. Infatti, a differenza di altre tecniche, permette di tener conto di più “livelli” di analisi consentendo di valutare l'effetto di proprietà di livello macro sugli atteggiamenti e comportamenti degli attori sociali.

¹⁷ In realtà, Raudenbush, Rowan e Kang (1991) hanno sviluppato una procedura per implementare un modello di misurazione in un modello multilivello composto da tre livelli. Tuttavia, in tale procedura i *factor loadings* non sono liberamente stimati ma devono essere specificati in precedenza (vedi anche Li, Duncan, Harmer, Acock e Stoolmiller 1998; Bauer 2003).

1988; Muthén 1989, 1990, 1994; Asparouhov e Muthen 2007; Mehta e Neale 2005; Rabe-Hesketh, Skrondal e Pickles 2004). L'obiettivo era quello di creare un approccio metodologico generale che fosse capace di dar conto contemporaneamente della struttura gerarchica dei dati, permettesse lo sviluppo di modelli di misurazione multilivello e consentisse di specificare modelli strutturali fra variabili (latenti ed osservate) appartenenti a diversi livelli di analisi.

I modelli di regressione multilivello e i modelli ad equazioni strutturali si sono sviluppati e sono evoluti in ambiti disciplinari diversi e si basano su presupposti concettuali e metodologici diversi (Depaoli e Clifton 2015; Mehta e Neale 2005). Tuttavia, numerosi autori ne hanno messo in luce i numerosi punti di connessione nonché la continuità che sussiste fra i due approcci (Bauer 2003; Curran 2003; Mehta e Neale 2005). Un esempio è il *latent growth model* (modello della curva di crescita latente), un modello longitudinale che implica la presenza di misure ripetute su una stessa variabile rilevata a diversi intervalli temporali (si veda Bollen e Curran 2006 per un approccio al LGM all'interno del paradigma dei modelli ad equazioni strutturali e Hox 2010 per un'applicazione dei modelli longitudinali all'interno del paradigma dell'analisi multilivello). I modelli longitudinali sono a tutti gli effetti modelli multilivello in quanto le misure ripetute possono essere pensate come unità di primo livello annidate (*nested*) all'interno di unità di secondo livello, ossia gli individui. Numerosi studi hanno mostrato come i classici modelli multilivello possono essere stimati attraverso un approccio SEM (Rovine e Molenaar 2000; Bauer, 2003; Curran 2003; Mehta e Neale 2005). Infatti, date alcune condizioni, quando abbiamo una struttura di dati longitudinale (misure ripetute su individui tale che le occorrenze delle variabili costituiscono il primo livello e gli individui il secondo) un modello ad equazioni strutturali è analiticamente equivalente ad un modello multilivello (Curran 2003; Mehta e Neale 2005). Per semplicità, consideriamo un modello con due livelli, ossia misure ripetute (livello-1) annidate in individui (livello-2) senza covariate al livello-2. L'equazione di primo livello sarà¹⁸:

$$Y_{ti} = \beta_{0i} + \beta_{1i}X_{1ti} + r_{ti}$$

Dove Y_{ti} è il valore della variabile y al tempo t per un individuo i ; x è la misura del tempo per un individuo i ; β_{0i} e β_{1i} sono, rispettivamente, l'intercetta e il coefficiente di regressione per un individuo i ; infine, r_{ti} è il consueto termine di errore che rappresenta il residuo riferito al singolo individuo. Le equazioni di secondo livello, non avendo covariate di livello-2 (in questo caso gli individui), sono semplificabili in:

$$\beta_{0i} = \gamma_{00} + u_{0i}$$

$$\beta_{1i} = \gamma_{10} + u_{1i}$$

Dove γ_{00} e γ_{10} rappresentano l'intercetta media e il coefficiente di regressione medio mentre i termini di errore, u_{0i} e u_{1i} , sono gli scarti degli individui da questi valori medi.

All'interno dei modelli ad equazioni strutturali il modello in parola assume la forma di un modello di analisi fattoriale confermativa. L'equazione del modello di misurazione assume la forma:

$$y_{it} = \eta_{\alpha i} + \lambda_t \eta_{\beta i} + \varepsilon_{it}$$

¹⁸ Nelle equazioni multilivello presenti nel paragrafo dedicato alla regressione multilivello sono stati usati i pedici i per indicare individui annidati in gruppi j , nelle equazioni che seguono saranno modificati in t e i rispettivamente per indicare le misure ripetute annidate in individui.

e il modello strutturale:

$$\eta_{\alpha i} = \mu_{\alpha} + \zeta_{\alpha i}$$

$$\eta_{\beta i} = \mu_{\beta} + \zeta_{\beta i}$$

Come notiamo dalle equazioni, i due modelli sono simili, dove i coefficienti *random* nelle equazioni presenti nel modello multilivello assumono la forma di variabili latenti nel modello ad equazioni strutturali. La differenza sta nel fatto che nel primo modello LGM la struttura multilivello è espressamente riconosciuta, mentre il modello LGM parametrizzato come SEM è essenzialmente un modello con un unico livello, in quanto i coefficienti *random* (ossia che variano a livello-2) sono trattati come variabili latenti. Per specificare un modello multilivello univariato (ossia un'unica variabile risposta) parametrizzato come un modello ad equazioni strutturali è necessario modificare l'unità d'analisi in modo che le unità di secondo livello diventino le unità di analisi e i valori delle variabili riferite agli individui siano considerate come variabili. In altre parole, gli individui costituiscono le variabili associate all'unità di analisi di secondo livello. Così come i modelli multilivello suddividono la varianza osservata in due componenti separate – la varianza *within* o all'interno dei gruppi e la varianza *between* o fra i gruppi – il modello di analisi fattoriale confermativa suddivide la varianza in una componente comune (ossia tra gli individui) e una specifica (ossia all'interno dell'individuo). Numerose indagini, in particolare Bauer (2003), Curran (2003) e Metha e Neale (2005), hanno mostrato come i classici modelli longitudinali (misure ripetute annidate in individui) e i classici modelli multilivello (per esempio studenti annidati in classi) possono essere riparametrizzati come modelli ad equazioni strutturali¹⁹.

2.4.4 I vantaggi dei modelli ad equazioni strutturali multilivello

Come detto, sono stati numerosi gli autori che hanno cercato di integrare l'approccio multilivello e quello dei modelli ad equazioni strutturali sviluppando un approccio che va sotto il nome di modelli ad equazioni strutturali multilivello (*multilevel structural equation modeling* abbreviato in MSEM o ML-SEM). I MSEM mettono insieme i vantaggi dei due approcci integrando i benefici derivanti dall'analisi multilivello – possibilità sia di analizzare dati *clusterizzati* producendo stime dei parametri non affette da *bias* sia di considerare nel modello gli effetti esercitati dalle variabili macro – e quelli derivanti dai SEM – possibilità di inserire nel modello variabili latenti, tener conto dell'errore di misurazione e testare ipotesi complesse (Selig, Card e Little 2008).

I modelli multilivello ad equazioni strutturali presentano numerosi vantaggi rispetto ai modelli multilivello classici. Infatti, i modelli multilivello ad equazioni strutturali costituiscono un approccio molto più flessibile dei modelli lineari gerarchici, ossia i modelli multilivello classici. Innanzitutto, permettono di incorporare nel modello costrutti latenti attraverso il ricorso ad indicatori multipli, laddove i modelli lineari gerarchici assumono invece la presenza di variabili perfettamente rilevate (Metha e Neale 2005; Preacher, Zhang e Zyphur 2011). Da un lato, la possibilità di integrare costrutti latenti all'interno del modello permette di sviluppare e testare teorie ed ipotesi più complesse – come

¹⁹ Dal momento che tale aspetto non costituisce oggetto specifico della tesi ma serve soltanto per mostrare la continuità che sussiste fra i modelli multilivello classici e i modelli ad equazioni strutturali, per ulteriori chiarimenti si rimanda a Rovine e Molenaar (1998, 2000, 2001), Bauer (2003), Curran (2003), Mehta e Neale (2005), Rabe-Hesketh, Skrondal e Pickles (2004).

detto, la maggioranza delle teorie presenti nelle scienze sociali postulano relazioni fra costrutti non direttamente osservabili. Dall'altro lato, la possibilità di integrare nel modello l'errore di misurazione (attraverso il ricorso ad indicatori multipli) permette di ottenere stime dei coefficienti strutturali maggiormente affidabili con tutti i risvolti positivi che ne conseguono in termini sostantivi riguardo le relazioni che sussistono fra le variabili e alla verifica di ipotesi complesse. Per fare alcuni esempi pratici, sappiamo che la stima del coefficiente di correlazione intraclasse viene influenzato dall'errore di misurazione delle variabili: la varianza del termine di errore a livello individuale tende a far aumentare il contributo della porzione di varianza attribuibile al livello-1 al computo complessivo del coefficiente di correlazione intraclasse. Quando nel modello si tiene in debito conto l'errore di misurazione, il rapporto fra varianza *within* e *between* risulterà depurato dal termine di errore ottenendo così stime maggiormente accurate del contributo apportato dal livello macro e quello viceversa prodotto dal livello micro (Muthén 1994). I MSEM permettono di specificare anche modelli di misurazione maggiormente affidabili in quanto la sorgente di errore ipotizzata non è univoca come avviene nei modelli classici di analisi fattoriale confermativa. Come vedremo, i MSEM ci permettono di scomporre le sorgenti di errore in una porzione attribuibile al livello micro e un'altra riconducibile al livello macro. Diversamente dai modelli multilivello classici, i MSEM sono una tecnica confermativa. Ciò permette di valutare l'adattamento generale del modello, ossia quanto il modello generale da noi teorizzato riproduce la matrice di varianze/covarianze presente nei dati (Ryu e West 2009; Hox 2013). Tale aspetto si configura come una possibilità in più per il ricercatore che ha l'opportunità di testare nel loro complesso teorie generali piuttosto che soffermarsi esclusivamente sull'influenza che una serie di predittori esercitano su una variabile indipendente.

I modelli ad equazioni strutturali multilivello permettono di specificare un'ampia serie di modelli teorici permettendo di specificare relazioni fra più costrutti posti a diversi livelli di analisi con la possibilità di integrare contemporaneamente nel modello sia variabili latenti che variabili direttamente osservate a qualsiasi livello della gerarchia di analisi (Muthén e Muthén 1998-2012). I MSEM permettono di integrare la classica *path analysis* con l'analisi fattoriale, tenendo contemporaneamente conto della struttura multilivello presente nei dati consentendo di ipotizzare relazioni fra variabili latenti/osservate a tutti i livelli della gerarchia (per un'applicazione pratica nell'ambito dei *cross-cultural research* si veda Cheung e Au 2005). Questo costituisce un notevole passo in avanti in quanto, generalmente, i modelli multilivello sono limitati nella misura in cui il modello può includere soltanto una variabile dipendente alla volta, laddove i MSEM permettono la specificazione di modelli con variabili dipendenti multiple (Metha e Neale 2005). Tale aspetto permette di sviluppare e testare modelli notevolmente più complessi attraverso la specificazione di catene "causali" che coinvolgano variabili latenti/osservate sia a livello micro che macro. Tale elasticità permette, inoltre, di scomporre gli effetti delle variabili in effetti diretti, indiretti e totale permettendo di testare ipotesi di mediazione o moderazione sia al primo che al secondo livello (Hox 2013, Preacher, Zyphur e Zhang 2010). In particolare, l'argomento riferibile all'analisi dei modelli di mediazione assume notevole importanza nei modelli ad equazioni strutturali multilivello. Sappiamo che le procedure classiche per testare un modello di mediazione (Baron e Kenny 1986) non sono più valide in presenza di strutture di dati ordinate gerarchicamente (Preacher, Zyphur e Zhang 2010). Numerosi autori (Krull e MacKinnon 1999, 2001; Pituch, Stapleton e Kang 2006), utilizzando un approccio riferibile ai modelli multilivello classici, hanno proposto metodi alternativi per analizzare gli effetti di mediazione in presenza di dati annidati gerarchicamente. Tuttavia, tali strategie soffrono di numerose limitazioni. In particolare, Preacher, Zyphur e Zhang (2010) individuano due problemi principali che sorgono con i modelli di mediazione stimati tramite un approccio multilivello classico. Innanzitutto, in un modello di mediazione multilivello l'effetto *within*

(livello-1) e l'effetto *between* (livello-2) vengono mischiati (*conflated*); per esempio, l'effetto *within* di una variabile mediatrice M su una variabile dipendente Y e l'effetto *between* della stessa variabile M su Y sono costretti ad essere uguali²⁰. Il limite maggiore, tuttavia, consiste nel fatto che, sebbene i modelli di mediazione multilivello classici possono includere una variabile indipendente a livello macro, tali modelli non possono gestire modelli con variabili mediatrici o variabili dipendenti collocate al secondo livello. Quindi i modelli con una variabile di livello-2 predetta da una variabile di primo o secondo livello (cioè modelli con disegno 1-1-2 o 1-2-2²¹) non possono essere stimati all'interno dei modelli lineari gerarchici (Preacher, Zyphur e Zhang 2010). I MSEM, al contrario, costituiscono uno strumento notevolmente più flessibile, permettendo di stimare modelli di mediazione molto più complessi contemplando la possibilità di inserire gli effetti di variabili di secondo livello (per esempio modelli dal disegno 2-1-2, 1-2-1, 1-2-2 e 1-1-2).

I modelli ad equazioni strutturali multilivello mostrano, come visto, notevoli vantaggi rispetto ai modelli lineari gerarchici. Ma non sono solo i vantaggi di tali modelli che ne giustificano la scelta, molte volte, come già in parte visto, si presentano come necessari. La loro mancata applicazione comporterebbe una serie di problematiche che inficerebbero pesantemente l'esito di un'indagine.

Assumendo una struttura di dati multilivello – come accade in tutte le indagini *cross-national* oppure nei campioni estratti con procedura a stadi o a grappoli, quindi molto più frequente di quanto si creda – il ricercatore, nell'analizzare la struttura fattoriale presente nei dati, può seguire tre strade distinte (Dyer, Hanges e Hall 2005; Bovaird 2007; Cheung e Au 2005). La “disaggregazione”, ossia condurre l'analisi sulla matrice di covarianza totale derivata dall'intero campione in modo da ignorare la struttura gerarchica dei dati e utilizzando semplicemente i dati individuali così come presenti nel campione. La “aggregazione”, ossia calcolare i punteggi medi aggregati per gruppo derivati dai singoli valori individuali – la classica analisi ecologica. Infine, l'analisi multilivello.

La disaggregazione produce stime degli errori standard che risultano sottostimati e i test statistici risulteranno in tal modo effetti da *bias* in modo che gli errori di 1 Tipo saranno maggiori del valore nominale (Bovaird 2007; Cheung e Au 2005; Raudenbush e Bryk 2002).

Il secondo approccio, invece, conduce ad una serie di problemi diversi. Innanzitutto, con i dati aggregati è più probabile che si commetta la cosiddetta “fallacia ecologica”, ossia un errore logico nel trarre conclusioni su relazioni individuali ottenute in realtà a livello aggregato (Robinson, 1950). Inoltre, i risultati ottenuti a livello aggregato ignorano totalmente potenziali relazioni significative che avvengono a livello individuale. Infine, i dati aggregati non sono una rappresentazione fedele dei dati a livello di gruppo. In termini più precisi, Muthén (1994) mostra come tali approcci producano risultati problematici. Quando viene analizzata la matrice di covarianza totale, sia l'adattamento della struttura fattoriale a livello di gruppo sia le stime dei *factor loadings* risulteranno affette da *bias* in quanto la struttura fattoriale è una combinazione della strutture fattoriali che operano al livello *between* e al livello *within* – in genere la struttura fattoriale è principalmente una funzione della struttura *within* (Muthén 1994). Inoltre, la matrice di covarianza a livello di gruppo (*between*) è una funzione non solo della struttura di covarianza appartenente al livello *between* ma anche della struttura di covarianza a livello *within* (Muthén 1994). Per risolvere tali problematiche che possono inficiare il risultato di un intero percorso di indagine l'approccio consigliato è quello riferibile all'analisi multilivello ad equazioni strutturali.

²⁰ Per ulteriori approfondimenti si rimanda a Preacher, Zyphur e Zhang 2010, Preacher, Zhang e Zyphur 2011, Zhang, Zyphur e Preacher 2009

²¹ Tale notazione riferita ai modelli di mediazione è stata suggerita da Krull e MacKinnon 2001.

2.4.5 La logica sottostante i modelli ad equazioni strutturali multilivello: la scomposizione della varianza e i diversi metodi di stima

In letteratura sono stati proposti essenzialmente due approcci ai modelli ad equazioni strutturali multilivello. L'approccio tradizionale (Muthén 1989, 1994; Muthén e Satorra 1989, 1995; Asparouhov e Muthén 2007) si basa sul principio di scomposizione della matrice di covarianza in due componenti ortogonali: la componente entro i gruppi (*within-groups*) e la componente tra i gruppi (*between-groups*). Rabe-Hesketh, Skrondal e Pickles (2004) hanno proposto un modello alternativo che non prevede la scomposizione della covarianza nelle due componenti separate²². Il primo approccio è stato denominato da Rabe-Hesketh, Skrondal e Zheng (2007, 2012) *within-between framework*, mentre il secondo approccio va sotto il nome di *generalized linear latent and mixed models* (GLLAMM) (Rabe-Hesketh, Skrondal e Pickles 2004; Skrondal e Rabe-Hesketh 2004; Rabe-Hesketh, Skrondal e Zheng 2007; Rabe-Hesketh, Skrondal e Zheng 2012)²³.

I MSEM assumono la presenza di una popolazione di individui divisa in gruppi (*nested*). Se non avessimo una tale struttura di dati (cioè individui annidati in unità di livello superiore) non avrebbe senso ricorrere ad un approccio multilivello. In questo caso, il vettore di una variabile osservata, indicato con Y_{ij} (i per gli individui e j per i gruppi), si può scomporre in una componente che varia tra i gruppi, $Y_B = \bar{Y}_j$, e una componente che varia entro i gruppi, $Y_w = Y_{ij} - \bar{Y}_j$ (Muthén 1989, 1994, si veda anche Hox 2010, Kaplan e Kreisman 2000, Cheung e Au 2005, Dyer, Hanges e Hall 2005). In altre parole, il valore totale osservato $Y_t = Y_{ij}$ viene sostituito con una componente di gruppo Y_B , che rappresenta il contributo delle unità di secondo livello, ed una componente individuale Y_w che rappresenta il contributo individuale al valore totale osservato (Kaplan e Kreisman 2000). Le componenti *between* e *within* hanno la proprietà di essere ortogonali ed additive (Hox 2010), ossia:

$$Y_t = Y_B + Y_w$$

Da ciò discende che la matrice di varianza e covarianza totale può essere scomposta a sua volta in due componenti ortogonali ed additive. Quindi, abbiamo una matrice di covarianza tra i gruppi, Σ_B , che è la matrice di covarianza della popolazione ottenuta come scarti delle medie di gruppo dalla media generale e una matrice di covarianza entro i gruppi, Σ_w , che rappresenta la matrice di covarianza della popolazione ottenuta come scarti dei valori individuali dalle medie di gruppo:

$$\Sigma_T = \Sigma_B + \Sigma_w.$$

Allo stesso modo possiamo scomporre le osservazioni campionarie ottenendo due componenti nuovamente ortogonali ed additive, ossia:

$$S_T = S_B + S_w.$$

Anche i termini di errore sono scomponibili in due componenti distinte, tale che:

²² Sarà posta maggiore attenzione sul primo approccio in quanto, oltre ad essere quello più diffuso attualmente, è stato anche quello a cui è stato fatto ricorso per l'analisi dei dati a nostra disposizione i cui risultati saranno presentati nel prossimo capitolo.

²³ Gli autori, contestualmente alle indagini dove hanno presentato il loro approccio, hanno rilasciato anche un programma statistico, nominato *gllamm*, che lavora in Stata.

$$\varepsilon_{ij} = \varepsilon_{Bj} + \varepsilon_{wij}$$

dove ε_{Bj} rappresenta la componente casuale che cattura l'effetto di gruppo e ε_{wij} è la componente casuale che cattura l'effetto individuale. Le due componenti, come le precedenti, sono fra loro ortogonali, $\text{cov}(\varepsilon_{Bj}, \varepsilon_{wij}) = 0$. Quindi anche la varianza residua può essere scomposta in una porzione riferita ai gruppi ed una riferita agli individui:

$$\text{Var}(\varepsilon_{ij}) = \Theta_B + \Theta_w.$$

Per stimare i parametri di un modello ad equazioni strutturali multilivello abbiamo bisogno delle stime di massima verosimiglianza della matrice di covarianza *between-groups* della popolazione (Σ_B) e della matrice di covarianza *within-groups* della popolazione (Σ_w). I dati presenti in un qualsiasi *dataset* si riferiscono alla matrice osservata *between-groups* (S_B) e alla matrice osservata *within-groups* (S_w). Ma, non possiamo semplicemente usare S_B come una stima di Σ_B e S_w come una stima di Σ_w (Muthén 1994, Hox 2010). Infatti, Muthén (1989) dimostra come la stima non affetta da *bias* della matrice di covarianza *within-groups* della popolazione (Σ_w) è data dalla cosiddetta matrice di covarianza *pooled within-groups*, S_{PW} , così come la matrice di covarianza scalata *between-groups*, S_B^* , è uno stimatore composito di massima verosimiglianza delle matrici di covarianza della popolazione *within* e *between*, ossia $S_B^* = \Sigma_w + c\Sigma_B$ (dove c è la numerosità dei gruppi). L'approccio proposto da Muthén (1989, 1994) prende il nome di approccio pseudobilanciato (McDonald 1994) o approccio ad informazione limitata (Hox 2010) o anche approccio MUML come denominato dallo stesso Muthén (1989, 1994). Tale approccio si differenzia a seconda della numerosità campionaria dei gruppi a seconda se tutte le unità di secondo livello hanno una numerosità campionaria interna uguale (gruppi con uguale numerosità) o viceversa i gruppi hanno diversa numerosità. Nel caso di gruppi bilanciati, abbiamo due matrici di covarianza campionarie: la matrice di covarianza *pooled within-groups*, S_{PW} , e la matrice di covarianza scalata *between-groups*, S_B^* .

La matrice di covarianza *pooled within-groups* si ottiene:

$$S_{PW} = \frac{\sum_j^G \sum_i^n (\mathbf{y}_{ij} - \bar{\mathbf{y}}_j)(\mathbf{y}_{ij} - \bar{\mathbf{y}}_j)'}{N - G}$$

Dal momento che la matrice campionaria pesata entro i gruppi (S_{PW}) è uno stimatore corretto della matrice di covarianza della popolazione entro i gruppi (Σ_w) si possono stimare i parametri della popolazione direttamente stimando il modello entro i gruppi attraverso la matrice S_{PW} . Mentre la matrice di covarianza scalata *between-groups* (S_B^*) si ottiene:

$$S_B^* = \frac{\sum_j^G n(\bar{\mathbf{y}} - \bar{\mathbf{y}}_j)(\bar{\mathbf{y}} - \bar{\mathbf{y}}_j)'}{G - 1}.$$

Ricapitolando, la matrice di covarianza campionaria *pooled within-groups* (S_{PW}) è la stima di massima verosimiglianza di Σ_w e la matrice di covarianza campionaria scalata *between-groups* (S_B^*) è la stima di massima verosimiglianza di $\Sigma_w + c\Sigma_B$.

Tale procedura si applica solo nel caso di gruppi bilanciati, ossia con la stessa numerosità campionaria. Nel caso di gruppi non bilanciati per la componente *within* rimane tutto invariato ma la porzione *between* del modello necessita di modifiche. Infatti, nel caso di gruppi con diversa numerosità

campionaria la matrice di covarianza campionaria scalata *between-groups* (S^*_B) stima una matrice diversa per ogni gruppo avente numerosità diversa (con stimatore FIML). Ciò implica la necessità di stimare un modello *between* separato per ogni gruppo con numerosità diversa determinando in tal modo una specificazione di una serie di modelli che renderebbero l'analisi complessa ed ingestibile in quanto in alcuni casi potrebbe accadere di avere gruppi con numerosità campionaria inferiore ai parametri della corrispondente matrice di covarianza (Hox 2010). La stima di massima verosimiglianza ad informazione completa (FIML), quindi, risulta essere problematica in questo caso. Per ovviare a ciò Muthén (1989) propone uno stimatore di massima verosimiglianza ad informazione limitata ignorando semplicemente il fatto che i gruppi sono sbilanciati (da qui il nome approccio pseudobilanciato). In tal modo risulta possibile calcolare una singola matrice di covarianza campionaria scalata *between-groups* (S^*_B), e non una per ogni gruppo avente numerosità diversa. Il modello proposto contiene uno stimatore ad hoc, ossia C^{*24} , che serve a scalare i parametri e che corrisponde alla numerosità campionaria media (ricordiamo che C in precedenza si riferiva alla numerosità dei gruppi che era uguale per ogni gruppo).

Si ottiene così uno stimatore di massima verosimiglianza ad informazione limitata denominato MUMML (acronimo di Muthén's Maximum Likelihood). Come detto, il metodo di stima si basa sul principio di ignorare il fatto che i gruppi siano internamente sbilanciati in termini di numerosità campionaria e sostituire il tutto con la media. Al fine della valutazione della correttezza dello stimatore il problema si pone, quindi, a livello *between* (cioè al secondo livello o livello dei gruppi). Studi di simulazione (Hox e Maas 2001; Hox, Maas, e Brinkhuis 2010) mettono infatti in luce come a livello *within* non sussiste nessun problema in tutte le condizioni simulate. A livello *between* invece, sebbene i *factor loadings* siano correttamente stimati, sia le varianze residue che gli errori standard risultano sottostimati inficiando i test di significatività. Yuan e Hayashi (2005) mostrano come gli errori standard e il test del chi-quadrato risultano corretti soltanto quando la numerosità campionaria di secondo livello tende all'infinito e il coefficiente di variazione della numerosità dei gruppi tende a zero. Sia gli studi di Hox e Maas (2001) e Hox, Maas, e Brinkhuis (2010) che lo studio di Yuan e Hayashi (2005) concordano nel ritenere che campioni con numerosità molto elevata non producono stime maggiormente accurate se ciò si accompagna ad una struttura campionaria non bilanciata fra i gruppi. Quindi, in presenza di campioni fortemente squilibrati in termini di numerosità campionaria l'approccio MUMML produce stime degli *standard errors* e test di significatività che risultano distorti e tale *bias* non si riduce anche se la numerosità campionaria viene elevata.

Muthén (1989, 1994) propose tale tecnica in un periodo in cui ancora non erano disponibili i *software* adatti specificamente per l'analisi dei MSEM; tale approccio aveva il vantaggio di poter essere applicato facendo ricorso ai *software* disponibili sul mercato ed adatti esclusivamente per stimare modelli ad equazioni strutturali semplici. Attraverso una procedura multi-gruppo, complessa e macchinosa, si riusciva a stimare modelli ad equazioni strutturali multilivello trattando i due livelli di analisi come due gruppi distinti. Tuttavia, come Hox (2010) fa notare, tale approccio risulta ormai datato e superato da altri che si basano su stimatori di massima verosimiglianza ad informazione completa (Asparouhov e Muthén 2003) o sulla funzione dei minimi quadrati ponderati – WLS – (Asparouhov e Muthén 2007). Tale approccio è stato presentato in quanto costituisce uno dei punti di partenza su cui indagini successive si sono basate permettendo lo sviluppo dei modelli ad equazioni strutturali multilivello. In letteratura sono stati proposti nuovi stimatori basati sul metodo di stima di massima verosimiglianza ad informazione completa – FIML – (Asparouhov e Muthén 2003; Liang e Bentler 2004; du Toit e du

²⁴ Per la formula si rimanda a Muthén (1989, 1994) o Hox (2002, 2010).

Toit 2008). Molti di questi stimatori sono basati sul cosiddetto *expectation-maximization algorithm* che permette di stimare i parametri del modello anche in presenza di dati mancanti. In particolare, Asparouhov e Muthén (2003) hanno proposto tre diversi metodi di stima di massima verosimiglianza basati sull'algoritmo EM. I tre stimatori, nominati ML, MLF e MLR (attualmente implementati nel *software* statistico *Mplus*) si caratterizzano per la particolare procedura attraverso la quale vengono calcolati gli *standar errors*. In particolare, lo stimatore ML produce stime di massima verosimiglianza con gli errori standard e il test del chi-quadrato calcolati in modo convenzionale; invece, lo stimatore MLR è uno stimatore robusto che produce stime di massima verosimiglianza con gli errori standard e il test del chi-quadrato che non assumono la normalità dei dati, inoltre i test del chi-quadrato sono calcolati con le correzioni Satorra-Bentler (1994) e Yuan-Bentler (2000) (Muthén e Muthén 1998-2012). Quando le variabili non sono distribuite normalmente, i modelli multilivello si possono ancora stimare attraverso uno stimatore di massima verosimiglianza ma l'intera procedura per il calcolo degli effetti casuali richiede un'integrazione numerica (Muthén e Muthén 1998-2012). Quando il modello include molti effetti casuali, l'integrazione numerica diventa eccessivamente pesante da un punto di vista computazionale rendendo praticamente impossibile la stima dei modelli (non è raro, soprattutto in presenza di variabili dipendenti binarie, categoriali o ordinali, che ben presto il modello superi le capacità computazionali dei comuni PC oggi in commercio). Per ovviare a tali problematiche, Asparouhov e Muthén (2007) hanno proposto un metodo di stima alternativo²⁵. Tale approccio risulta particolarmente utile quando si è in presenza di variabili dipendenti binarie, categoriali o ordinali (difatti tale stimatore rappresenta il *default* nel *software* statistico *Mplus* quando si analizzano modelli che includono variabili binarie, categoriali o ordinali). Il metodo di stima, basato sui minimi quadrati ponderati – WLS – (Muthén 1984), è utile per stimare modelli ad equazioni strutturali multilivello con variabili osservate binarie, categoriali, ordinali, censurate e continue permettendo anche la combinazione fra i diversi tipi di variabili. Tale metodo (implementato in *Mplus* con lo stimatore WLSM o WLSMV definito “stimatore dei minimi quadrati diagonali”) risulta sufficientemente generale da poter stimare un elevato numero di modelli multilivello con la possibilità di includere variabili esplicative sia al primo che al secondo livello; inoltre, fornisce la possibilità di includere nel modello le intercette casuali ma non permette ai coefficienti di regressione di variare al secondo livello. Tale limite, tuttavia, è bilanciato dai numerosi vantaggi consistenti nella possibilità di stimare modelli complessi con tempi di calcolo ragionevoli o modelli che altri metodi di stima non permetterebbero nemmeno di stimare. Come suggerisce Hox (2010) e Hox, Maas, e Brinkhuis (2010) sia il metodo che ricorre alle stime di massima verosimiglianza che i minimi quadrati ponderati producono stime dei parametri che sono maggiormente accurati del metodo MUML prima esposto e che le stime prodotte dagli stimatori ML e WLS sono equivalenti. Hox (2010) conclude sostenendo che, per stimare un modello, gli stimatori di massima verosimiglianza sono la scelta da fare in prima battuta, ma se le condizioni dell'indagine rendono impraticabile tale opzione (per esempio per l'eccessivo tempo di calcolo richiesto) il metodo WLS è una alternativa ugualmente valida.

²⁵ Tale approccio sarà ulteriormente approfondito più avanti in quanto costituisce il metodo a cui è stato fatto ricorso per l'analisi dei dati.

2.4.6 Il modello di analisi fattoriale confermativa multilivello

I modelli ad equazioni strutturali multilivello sono una generalizzazione dei classici modelli ad equazioni strutturali. Uno dei modelli più importanti ed utili nell'ambito delle scienze sociali è il modello di analisi fattoriale confermativa che permette al ricercatore di misurare concetti complessi e non direttamente osservabili attraverso il ricorso ad un insieme di indicatori valutandone l'affidabilità, la validità convergente e discriminante (Bollen 1989).

Nell'ambito dei modelli ad equazioni strutturali multilivello, allo stesso modo, è stato sviluppato il modello di analisi fattoriale confermativa multilivello – MCFA – (Mehta e Neale 2005, Hox 2010) che svolge le stesse funzioni di quello precedente permettendo in aggiunta un'analisi più raffinata da un lato e l'individuazione di fattori/costrutti complessi che risiedono ad un livello superiore a quello individuale (per esempio individuando fattori attribuibili epistemologicamente ad aggregati territoriali come il benessere di una comunità, la sicurezza di un quartiere, la coesione di un gruppo sociale). La MCFA consente di applicare un'analisi fattoriale confermativa in presenza di una struttura di dati ordinati gerarchicamente. Abbiamo visto all'inizio del paragrafo proprio in riferimento all'analisi fattoriale confermativa come ignorare la struttura annidata dei dati conduca a risultati che risulterebbero gravemente affetti da *bias* (Julian 2001, Pornprasertmanit, Lee e Preacher 2014). Quando abbiamo a che fare con un campione annidato gerarchicamente, anche se l'interesse è esclusivamente a livello individuale, in nessun modo si dovrebbe ignorare la struttura di secondo livello entro cui gli individui sono *clusterizzati*. Come detto, se si analizzasse soltanto la matrice di covarianza totale, la struttura individuata a livello individuale è comunque frutto anche della struttura del livello aggregato (Muthén 1994). Quindi, anche se l'obiettivo dell'indagine contempla esclusivamente il piano individuale, in ogni caso si dovrebbe specificare una struttura fattoriale a livello aggregato (postulando che ve ne sia una). In questo caso, tutte le stime individuate a livello individuale rappresenterebbero gli effetti individuali depurati dall'influenza delle unità di secondo livello (Pornprasertmanit, Lee e Preacher 2014). Ignorando tale aspetto, le stime che si otterrebbero applicando una semplice analisi fattoriale sarebbero una combinazione spuria di effetti individuali ed effetti di contesto. Altra caratteristica fondamentale della MCFA è che permette l'individuazione simultanea di fattori posti a livello individuale e a livello aggregato che possono definirsi isomorfi fra di loro. In questo caso, il livello aggregato assume significato sostantivo in sé rappresentando, insieme al livello individuale, il fulcro dell'analisi. Il concetto di isomorfismo sta ad indicare il parallelismo che sussiste nella struttura o nelle funzioni che svolgono due costrutti a differenti livelli di analisi (per l'isomorfismo si veda van de Vijver, van Hemert e Poortinga 2008, Adamopoulos 2008 e Fontaine 2008; per le diverse relazioni funzionali che possono sussistere fra due costrutti a diversi livelli di analisi si veda Chan 1998). Empiricamente, due concetti a diversi livelli di analisi sono isomorfi quando mostrano la stessa struttura fattoriale e le relazioni fra indicatori e costrutti di primo e secondo livello mostrano la stessa intensità. In tal caso, esiste un parallelismo fra il costrutto presente al primo livello e quello presente al secondo livello indicando che il costrutto in parola opera allo stesso modo sia a livello aggregato che a livello individuale.

I modelli di analisi fattoriale confermativa multilivello permettono di definire simultaneamente un modello di misurazione a livello individuale attraverso il ricorso ad indicatori di primo livello e un modello di misurazione a livello aggregato per mezzo di indicatori di secondo livello, permettendo alle intercette degli indicatori di primo livello di variare al secondo livello (*random intercepts*) (Muthén 1994; Hox 2010). I modelli di MCFA scompongono la matrice di covarianza totale in due matrici indipendenti ed ortogonali, cioè la matrice di covarianza fra i gruppi e la matrice pesata entro i gruppi.

Utilizzando queste due matrici si riesce ad analizzare la struttura fattoriale simultaneamente a ciascuno dei livelli di analisi (Muthén 1994; Hox 2010; Cheung and Au 2005).

Il modello di analisi fattoriale confermativa a due livelli può essere scritto come segue (si veda anche Muthén 1991; Muthén 1994; Mehta and Neale 2005; Hox 2010; Selig, Card e Little 2008):

$$y_{ij} = \mu_j + \lambda_w \eta_{wij} + \varepsilon_{wij}$$

$$\mu_j = \mu + \lambda_b \eta_{bj} + \varepsilon_{bj}$$

La prima equazione rappresenta la porzione *within* (livello individuale) del modello; la seconda equazione descrive la porzione *between* (livello aggregato) del modello. Le due equazioni sono collegate tramite le intercette *random* indicate con μ_j . Le intercette specifiche delle unità di secondo livello (μ_j) sono allo stesso tempo variabili dipendenti nella porzione *between* dell'intero modello. Quindi, le intercette degli indicatori di primo livello riferite a ciascun unità di secondo livello sono variabili *random* al secondo livello – possono variare all'interno delle unità di secondo livello. Il fattore latente di secondo livello (η_{bj}) spiega la varianza delle intercette specifiche delle unità di secondo livello; la varianza non spiegata dal fattore di secondo livello, la cosiddetta varianza residua, è contenuta nel termine di errore di secondo livello (ε_{bj}).

2.4.7 Il modello MSEM completo: modello di misurazione e modello strutturale

Nella fattispecie si presenterà l'approccio riferibile a Asparouhov e Muthén (2007) in quanto costituisce il metodo a cui è stato fatto ricorso per analizzare i dati a nostra disposizione (si veda cap.3). Altri modelli sono stati proposti (per alcuni esempi si veda Rabe-Hesketh, Skrondal e Pickles 2004; Rabe-Hesketh, Skrondal e Pickles 2007; Rabe-Hesketh, Skrondal e Zheng 2012; Mehta e Neale 2005; Depaoli e Clifton 2015; Preacher, Zyphur e Zhang 2010).

In presenza di variabili non distribuite normalmente, i MSEM possono essere comunque stimati ricorrendo a stime di massima verosimiglianza, ma gli effetti casuali devono essere integrati attraverso una procedura numerica estremamente pesante da un punto di vista computazionale rendendo molte volte non stimabili modelli anche solo lievemente complessi. Tale problema si acuisce in presenza di variabili categoriali o ordinali creando ulteriori problemi di convergenza e allungamento ulteriore dei tempi di calcolo. Per ovviare a ciò, Asparouhov e Muthén (2007) hanno proposto un metodo basato sui minimi quadrati ponderati (*weighted least squares*) particolarmente utile per stimare modelli in cui gli indicatori sono costituiti (non solo) da variabili categoriali. In particolare, il modello proposto è sufficientemente generale da poter stimare modelli ad equazioni strutturali a due livelli con variabili osservate binarie, categoriali, ordinali, censurate, continue e tutte le combinazioni fra queste variabili. Lo stimatore WLS, proposto da Muthén per la prima volta nel 1984 (Muthén 1984) ed implementato nell'ultima versione del *software* Mplus come stimatore WLSMV, permette di stimare modelli con un numero qualsiasi di effetti casuali non soffrendo delle limitazioni viste in precedenza (tempi di calcolo eccessivi, problemi di convergenza).

Il modello considera tre tipi di variabili dipendenti: categoriche (che include anche le politomiche ordinali e le dicotomiche), censurate e continue. Il modello considera una variabile latente sottostante distribuita normalmente per ciascuna variabile osservata indicata con Y_{pij}^* . Tale variabile viene definita diversamente in base al tipo di variabile osservata Y_{pij} a cui è riferita. Nel caso di variabili continue

distribuite normalmente la variabile latente e la variabile osservata coincidono, $Y_{pij}^* = Y_{pij}$. In presenza di variabili categoriche (comprese ordinali politomiche e dicotomiche), la variabile latente sottostante è definita attraverso una serie di parametri soglia, τ_{pk} , tale che:

$$Y_{pij} = k \Leftrightarrow \tau_{pk} < Y_{pij}^* < \tau_{pk+1}$$

Dove K ($k = 0, 1, \dots, k-1$) rappresenta le modalità della variabile categoriale e i parametri soglia sono $-\infty = \tau_0 < \tau_0 < \tau_1 < \tau_2 < \dots < \tau_{k-1} < \tau_k = +\infty$.

In sintesi, per ogni variabile categoriale indipendente viene definita una variabile latente continua sottostante alla variabile categoriale che ne rappresenta soltanto una “manifestazione”.

I modelli ad equazioni strutturali multilivello, esattamente come i modelli ad equazioni strutturali classici, sono composti da due modelli, quello di misura e quello strutturale. Innanzitutto, la variabile latente Y_{pij}^* è scomponibile in due parti additive e ortogonali tale che $Y_{pij}^* = Y_{wpij} + Y_{bpj}$, dove Y_{wpij} rappresenta gli effetti individuali, ossia le differenze fra individui all'interno di un gruppo e Y_{bpj} racchiude gli effetti di gruppo (intercette *random*), ossia le differenze fra le unità di secondo livello.

Il modello di misurazione e il modello strutturale di primo livello possono essere scritti come:

$$Y_{wij} = \Lambda_w \eta_{wij} + \varepsilon_{wij}$$

$$\eta_{wij} = B_w \eta_{wij} + \Gamma_w X_{wij} + \zeta_{wij}$$

Nel modello di misurazione, Λ rappresenta la matrice dei coefficienti fattoriali mentre ε_{wij} costituisce il vettore dei residui. Nel modello strutturale, B sta ad indicare la matrice dei coefficienti di primo livello riguardanti le variabili latenti e Γ la matrice dei coefficienti fra le variabili latenti e le variabili esplicative osservate X con ζ_{wij} che rappresenta il termine di errore di primo livello.

Allo stesso modo possiamo scrivere il modello di misurazione e il modello strutturale di secondo livello:

$$Y_{Bj} = v_B + \Lambda_B \eta_{Bj} + \varepsilon_{Bj}$$

$$\eta_{Bj} = \alpha_B + B_B \eta_{Bj} + \Gamma_B X_{Bij} + \zeta_{Bj}$$

Come riportato in Asparouhov e Muthén (2007), i parametri da stimare si riferiscono alle matrici Λ_w , Γ_w , B_w , Λ_B , Γ_B e B_B e ai vettori v_B , α_B e τ_{pk} (questi ultimi ricordiamo sono i parametri soglia). Non tutti i parametri appena descritti devono essere stimati, ma possono essere fissati a zero per ottenere diversi modelli strutturali. Inoltre, per problemi di identificazione, soltanto in riferimento alla variabili categoriali la varianza dei residui di primo livello, ossia ε_{wij} , è fissata ad 1; quindi, tutti i coefficienti riferiti a tali variabili avranno la caratteristica di essere standardizzati.

Stimare un modello che contiene molteplici variabili categoriali dipendenti attraverso uno stimatore di massima verosimiglianza molte volte risulta impossibile in quanto ogni variabile categoriale (più gli effetti casuali) richiede una procedura di integrazione numerica²⁶ generando tre ordini di problemi: i tempi di stima del modello risultano smisurati, si generano numerosi problemi di convergenza e, infine, date le numerose procedure di integrazione numerica il carico computazionale risulta ingestibile dai comuni *personal computer* attualmente in commercio eccedendo la memoria disponibile e rendendo

²⁶ Nello specifico, la stima univariata di una variabile categoriale richiede una dimensione di integrazione numerica, la stima bivariata riferita a due variabili categoriali utilizza due dimensioni di integrazione numerica e, infine, la stima bivariata riguardante una variabile continua ed una categoriale richiede una dimensione di integrazione numerica.

di fatto non stimabile il modello di nostro interesse. Lo stimatore WLS permette di superare tutti i limiti appena esposti rendendo di fatto stimabili tutta una serie di modelli che altrimenti non lo sarebbero. Sono stati effettuati studi di simulazione (Asparouhov e Muthén 2007) per comparare le prestazioni dello stimatore WLSMV con quelle dello stimatore di massima verosimiglianza (ML). Nella fattispecie, il modello di riferimento era costituito da un modello di analisi fattoriale confermativo multilivello con variabili ordinali e lo stesso modello stimato attraverso variabili continue. Con le variabili continue il tasso di convergenza è stato del 100% sia per lo stimatore WLSMV che per lo stimatore ML, invece, in presenza delle variabili ordinali, lo stimatore WLSMV ha conservato lo stesso tasso di convergenza mentre è sceso al 47% per lo stimatore ML. In riferimento all'efficienza delle stime, in presenza di variabili continue entrambe le stime ottenute con lo stimatore WLSMV e con lo stimatore ML non risultano distorte, mentre le stime ottenute attraverso lo stimatore WLSMV risultano essere più efficienti e meno distorte di quelle ottenute tramite stimatore ML in presenza di variabili osservate dipendenti ordinali (oltre a Asparouhov e Muthén 2007 si veda anche Hox, Maas, e Brinkhuis 2010).

2.4.8 Le strategie di analisi

In riferimento alle strategie di analisi in letteratura sono individuabili tre procedure distinte (Selig, Card e Little 2008). L'approccio di Muthén (1994, si veda anche Dyer, Hanges e Hall 2005 e Cheung e Au 2005), composto da quattro procedure consequenziali, prevede la suddivisione della matrice di covarianza totale in due parti distinte: la matrice di covarianza pesata entro i gruppi e la matrice di covarianza fra i gruppi. Il primo passaggio prevede l'analisi della matrice di covarianza totale ignorando, in tal modo, la struttura gerarchica dei dati. Tale modello risulterà certamente non corretto in quanto ignora la natura multilivello dei dati, ma serve a fornire al ricercatore una prima approssimazione di come potrebbe essere la struttura fattoriale e le relazioni all'interno del campione. Il secondo passaggio è utile ad evidenziare se nei dati a disposizione del ricercatore c'è effettivamente una variazione anche a livello di gruppo (non solo individuale) rendendo necessario il ricorso ad un approccio multilivello. Per valutare l'ammontare di varianza a livello di gruppo il ricercatore può far ricorso al coefficiente di correlazione intraclasse (Muthén 1994, Hox 2010, Dyer, Hanges e Hall 2005). Nel terzo *step* si procede ad analizzare solo il livello individuale facendo ricorso alla matrice di covarianza pesata entro i gruppi ignorando in tal modo del tutto il livello riferibile ai gruppi. Una volta individuata la struttura fattoriale e le relazioni fra le variabili (tale struttura così come le relazioni sono totalmente depurate dagli effetti delle unità di secondo livello) si passa al quarto *step* nel quale si procede ad analizzare le unità di secondo livello individuando una struttura fattoriale di secondo livello facendo ricorso alla matrice di covarianza fra i gruppi.

La strategia proposta da Hox (2002, 2010) risulta essere meno macchinosa e più intuitiva. Si basa sulla stima in serie di un insieme di modelli (*benchmark models*) sui quali vengono testate alcune ipotesi e successivamente confrontati fra di loro. Innanzitutto si stima il cosiddetto "modello nullo" nel quale viene stimata soltanto la struttura di primo livello assumendo in tal modo l'assenza di una struttura di secondo livello. Se tale modello mostra un buon adattamento ai dati vuol dire allora che non vi è nessuna variazione significativa al secondo livello (non vi è nessun effetto di gruppo). Se il modello nullo presenta uno scarso adattamento ai dati allora si procede a stimare il cosiddetto "*independence model*" nel quale vengono stimate, oltre al modello di primo livello, le varianze delle variabili osservate al secondo livello mentre le covarianze fra tali variabili vengono fissate a zero (non sono stimate). Se tale modello mostra un buon adattamento ai dati vuol dire che vi è una variazione all'interno delle

unità di secondo livello, ma al tempo stesso non è individuabile una struttura che sintetizzi tale variazione; in altri termini, tra le variabili non vi è covariazione al secondo livello. Nel terzo passaggio, Hox (2002, 2010) suggerisce di stimare il cosiddetto “modello saturo” nel quale i parametri riferiti alle covarianze delle variabili al secondo livello – prima fissate a zero – vengono liberamente stimate. Tale modello, dal momento che mostra il miglior adattamento possibile al secondo livello (è un modello saturo al livello *between*), serve come modello comparativo per giudicare la significatività di eventuali altre restrizioni da apportate ai parametri del modello di secondo livello. Ovviamente, come Hox fa notare (2002, 2010), le modifiche al modello devono essere teoricamente orientate, altrimenti si rischia di perdersi fra innumerevoli modelli alternativi ugualmente validi.

Il terzo approccio, riconducibile a Mehta e Neale (2005), prevede, come primo passo, la stima di un modello ad intercette variabili univariato al fine di ottenere i valori del coefficiente di correlazione intraclasse che ci suggerisce l'ammontare della varianza di secondo livello in rapporto alla varianza totale. Tale passaggio è utile per valutare se è necessario intraprendere un'analisi multilivello o, piuttosto, è sufficiente applicare un approccio classico, meno dispendioso in termini di risorse cognitive e temporali. Successivamente, si procede a stimare un modello ad intercette variabili multivariato con lo scopo di ottenere la matrice di covarianza fra i gruppi e la matrice di covarianza entro i gruppi le quali ci mostrano la struttura di covarianza presente nei nostri dati. Infine, come ultimo passaggio, gli autori suggeriscono di stimare un modello di analisi fattoriale confermativa sul quale è possibile successivamente testare tutta una serie di ipotesi aggiuntive. Riassumendo, l'approccio di Muthén in quattro fasi (1994) risulta ormai superato in quanto prevede la suddivisione e il calcolo delle matrici *between* e *within* ponendo inoltre un forte accento sulle fasi preliminari dell'analisi. C'è da dire, in ogni caso, che tale approccio si contraddistingueva come pioneristico nel momento in cui fu presentato e si poneva come unica soluzione per stimare un modello ad equazioni strutturali multilivello facendo ricorso ai *software* classici presenti all'epoca sul mercato i quali non prevedevano una specifica funzione per stimare i MSEM. La strategia di Muthén, si imponeva come procedura necessaria al fine di stimare un modello attraverso lo stimatore introdotto dallo stesso Muthén (lo stimatore MUML di cui abbiamo parlato in precedenza), tuttavia, allo stato attuale, con l'introduzione della stima di massima verosimiglianza ad informazione completa (*Full Information Maximum Likelihood*, FIML) tale procedura di scomposizione della matrice di covarianza totale non è più necessaria (Selig, Card e Little 2008). L'approccio di Mehta e Neale (2005) è quello più elegante in quanto mostra chiaramente la continuità che sussiste fra i modelli di regressione multilivello e i modelli multilivello ad equazioni strutturali (Selig, Card e Little 2008) ma è anche quello che richiede il maggior dispendio di risorse temporali in quanto richiede il ricorso a *software* differenti a seconda delle diverse fasi di analisi (a tal proposito si veda Mehta e Neale 2005). Secondo Selig, Card e Little (2008), l'approccio riconducibile ad Hox (2002, 2010) risulta essere quello maggiormente chiaro e semplice da implementare rispetto agli altri due.

Per concludere si può sostenere che i modelli ad equazioni strutturali multilivello costituiscono un nuovo strumento a disposizione dei ricercatori molto performante che sicuramente, quando la sua diffusione avverrà su larga scala, fornirà nuova linfa agli studi riconducibili alle scienze sociali. Contrariamente a quanto semplicisticamente si possa pensare, tale sviluppo non sarà confinato all'ambito riferibile meramente all'analisi dei dati, ma avrà riscontri anche sugli impianti teorici e sulla raffinatezza delle ipotesi formulabili. Ora che vi sono strumenti adatti a testare ipotesi complesse multilivello, le teorie sociali dovrebbero abbracciare una visione più ampia che contempli più livelli di analisi simultaneamente e metta in luce le relazioni di continuo scambio che sussistono fra l'individuo e il suo ambiente.

3. Operativizzazione e misurazione del concetto di coesione sociale: un modello ad equazioni strutturali multilivello

3.1 Una proposta di definizione operativa della coesione sociale

3.1.1 Considerazioni preliminari

Sulla base dello schema proposto nel capitolo 1, fornito allo scopo di facilitare le fasi di definizione operativa e di misurazione del concetto di coesione sociale e tenendo in debito conto i contributi presenti in letteratura relativi al concetto in esame, si è proceduto alla formulazione di una nostra proposta di operativizzazione della coesione sociale.

In letteratura, come mostrato, vi sono stati numerosi tentativi di operativizzare il concetto in esame ma molto spesso tali contributi hanno avuto il grave limite di confondere gli aspetti che si configurano come cause/effetti della coesione sociale con quelli che ne rappresentano le dimensioni costitutive (Chan et al. 2006). Soltanto per fornire qualche esempio a riguardo possiamo notare come Berger-Schmitt (2000) considera gli aspetti legati alle disuguaglianze come una dimensione fondamentale della coesione sociale (si veda anche Bernard 1999). L'ineguaglianza, all'interno della nostra prospettiva concettuale, si pone piuttosto come un fattore esterno, una causa, che può incidere sui livelli di coesione di uno Stato ma non per questo ne costituisce una sua dimensione fondamentale. Gli aspetti relativi all'esclusione o inclusione all'interno della sfera economica (per esempio il mercato del lavoro) così come gli elementi relativi alle disparità nei livelli di reddito percepito costituiscono cause (o effetti) della coesione sociale e, in quanto tali, dovrebbero costituire ipotesi da testare empiricamente piuttosto che considerati aprioristicamente come dimensioni costituenti.

In base alla prospettiva concettuale seguita, dalla nostra proposta di definizione operativa sono stati così esclusi tutti quei fattori che potevano configurarsi come cause della coesione evitando in tal modo di confondere le dimensioni che costituiscono un concetto con le cause che possono generarlo.

Dalla definizione operativa proposta sono stati eliminati inoltre tutti gli elementi afferenti alle dimensioni relative alla condivisione di valori e al sentimento di una identità comune. Numerosi autori (Jenson 1998, Bernard 1999) hanno considerato tale elemento una dimensione della coesione sociale, ma la scelta di non tenerne conto nella definizione operativa è stata effettuata sulla base di diverse considerazioni teoriche. Innanzitutto, non è assolutamente chiaro e scontato quali siano i valori che le persone dovrebbero condividere. Inoltre, non risulta essere di facile risoluzione il quesito relativo a quali soggetti o istituzioni spetta l'ultima parola su quali siano i valori meritevoli di essere condivisi. Infine, i valori assumono molte volte contorni sfumati e la loro intensione concettuale risulta essere poco definita e non chiara. Tutti gli individui hanno ben presenti i principi generali che devono essere rispettati e seguiti all'interno della società in cui si vive, ma spesso tali valori risultano essere eccessivamente generici per cui nessun individuo socializzato ammette di non riconoscerli come tali (ad esempio la democrazia). Viceversa, tali valori possono risultare eccessivamente specifici, tanto da diventare opinabili (ad esempio è preferibile la democrazia diretta o la democrazia rappresentativa?). In quest'ultimo caso, si corre il rischio che tutte le visioni siano in un modo o nell'altro legittime, si ritorna così al secondo problema: chi decide quale valore (o visione) sia "più legittimo"?

Inoltre, alcuni aspetti relativi ai valori che implicano un senso di identità condivisa risultano essere ambigui. Alcuni indicatori del senso di appartenenza inseriti nei questionari sotto forma di domande come “nonostante i suoi molti difetti questo Paese è ancora la mia casa” o “quanto è forte il tuo senso di appartenenza al tuo Paese” potrebbero essere indicatori poco validi. Più precisamente, la loro parte estranea (cioè la parte non indicante) potrebbe essere molto ampia indicando un concetto diverso da quello che dovrebbe rilevare. Per esempio, le risposte alle domande riportate precedentemente potrebbero indicare non un senso di appartenenza ma, piuttosto, un sentimento nazionalista, che non è il concetto di coesione che a noi interessa rilevare. Tali distorsioni possono essere attenuate inserendo domande di controllo relative al sentimento di apertura e tolleranza nei confronti di individui provenienti da altre Nazioni, ma rimane comunque poco chiaro cosa stiamo effettivamente rilevando.

Un ulteriore aspetto da considerare, che ci pone in opposizione alla formulazione proposta da Chan et al. (2006), riguarda la dimensione afferente al rispetto delle diversità e al sentimento di tolleranza. Chan et al. (2006) propone di eliminare dalla concettualizzazione della coesione sociale l'aspetto relativo alla tolleranza verso le diversità, non considerando tale elemento una dimensione costituente della coesione. A nostro parere, questa scelta è poco avveduta se si vuole tener conto della società multietnica e multiculturale in cui viviamo – aspetto la cui rilevanza tenderà a crescere in un futuro prossimo. Ignorare tale elemento non risolve di certo la questione. In una società multiculturale se i cittadini non mostrano la disponibilità a rispettare le differenze altrui e a riconoscere gli altri individui come esseri umani portatori di differenze e specificità l'ordine sociale potrebbe essere messo a rischio. Vi è un ulteriore aspetto a riguardo che ci ha fatto propendere per l'inserimento di tale dimensione nella concettualizzazione della coesione sociale. Esistono due tipi di legami: *bridging ties*, ossia legami positivi che connettono individui provenienti da gruppi differenti e *bonding ties*, ovvero legami che sussistono esclusivamente fra individui appartenenti allo stesso gruppo sociale (Putnam 2001). Se si è interessati a rilevare la coesione all'interno di una società allora non bisognerebbe prendere in considerazione esclusivamente gli aspetti legati alle relazioni sociali ma analizzare anche se tali relazioni siano “segregate” all'interno di un gruppo senza nessun legame esterno. Infatti, la coesione sociale costituisce un concetto duplice. Se un gruppo sociale mostra un'eccessiva coesione intra-gruppo potrebbe mostrare la tendenza ad escludere altri gruppi sociali generando il paradosso “can social cohesion be a threat to social cohesion?” in quanto “cohesive communities can suffer from too much ‘bonding’. One can be made only too aware that one is ‘not from the neighbourhood’ and therefore an object of suspicion, that one is not ‘from the old gang’ and therefore an outsider” (Jenson 1998, 36). Per tali ragioni, è stato ritenuto opportuno inserire all'interno della concettualizzazione della coesione sociale la dimensione afferente al rispetto e all'accettazione delle diversità.

3.1.2 La definizione operativa della coesione sociale

La definizione operativa proposta cerca di tenere in debito conto sia i contributi teorici forniti dai diversi autori illustrati nel capitolo 1 sia lo schema concettuale della coesione sociale proposto (cfr. cap.1). Pertanto, si basa su una serie di aspetti che ne costituiscono, a nostro avviso, i pilastri portanti. Innanzitutto, riformulando la definizione di Chan et al. (2006), intendiamo la coesione sociale come “una condizione riguardante le interazioni fra membri di una società, caratterizzati da una serie di atteggiamenti, norme e dalla loro manifestazione comportamentale che include la forza dei legami interpersonali nelle reti sociali primarie e secondarie, fiducia, partecipazione e riconoscimento di un

insieme di regole, delle istituzioni democratiche e della società nel suo insieme". Tale definizione mette in risalto come la coesione sia considerata una condizione e non un processo evitando in tal modo di confondere i mezzi e i fini (come accadeva in alcune proposte presentate nel cap. 1). In base ad una prospettiva riferibile all'individualismo metodologico, la definizione considera le relazioni e i legami positivi che sussistono all'interno della società. Viene espressamente reso manifesto, inoltre, che gli attori sociali debbano riconoscere le norme, le regole e la legittimità delle istituzioni democratiche al fine di escludere da tale definizione i gruppi ben coesi al loro interno ma finalizzati ad attività criminali. In secondo luogo, rifacendoci ai suggerimenti di Whelan and Maître (2005), la definizione operativa prende in considerazione tre ambiti di analisi: micro, meso e macro. Il primo si riferisce alle relazioni che sussistono fra individui, l'aspetto meso è relativo alle relazioni vigenti fra individui e gruppi (come l'individuo si raffronta con diversi gruppi), infine, l'ambito macro riguarda le relazioni sussistenti fra individuo e società nel suo complesso.

Inoltre, seguendo Bollen and Hoyle (2001), la definizione operativa tiene in debito conto due prospettive di analisi all'interno degli ambiti sopra illustrati. La prospettiva soggettiva è riferibile agli elementi riguardanti la sfera degli atteggiamenti, mentre la prospettiva oggettiva considera gli aspetti riferibili all'ambito dei comportamenti.

Sempre seguendo lo schema da noi proposto, la definizione operativa prende in considerazione due livelli di analisi: quello riferibile agli individui e quello relativo allo Stato valutandone l'effetto aggregato sugli atteggiamenti e comportamenti individuali. Come unità di analisi di secondo livello è stato scelto lo Stato in quanto ancora oggi costituisce una delle principali entità che contribuisce a formare l'identità del soggetto trasmettendone valori, tradizioni, regole dell'agire sociale ed è dotato di apposite istituzioni funzionali alla trasmissione intergenerazionale di tale identità.

Tenendo in debito conto i tre ambiti di analisi, le due prospettive e i due livelli di analisi abbiamo definito le dimensioni costituenti della coesione sociale come riportato nella tab.3.1.

Tab.3.1 – Dimensioni della coesione sociale

	soggettiva	oggettiva	Livello
Micro	Fiducia interpersonale	Densità delle relazioni sociali	Individuale
	Supporto sociale		
Meso	Apertura	Partecipazione	<i>Non considerato</i>
Macro	Fiducia nelle istituzioni	Legittimità delle istituzioni	Paese

L'ambito micro della coesione sociale include le dimensioni riguardanti la "fiducia interpersonale", il "supporto sociale" e la "densità delle relazioni sociali". La sfera meso contiene invece la dimensione riguardante l'accettazione e l'apertura nei confronti di soggetti stranieri e "diversi" e la dimensione afferente alla partecipazione politica e sociale. Nell'ambito macro vi sono due dimensioni strettamente connesse: la "fiducia nelle istituzioni" e la "legittimità di tali istituzioni".

Abbiamo eliminato dalla definizione operativa la dimensione relativa alla condivisione di valori per le motivazioni precedentemente esposte. Un'ulteriore motivazione, in modo ancor più rilevante e teoricamente orientato, che ha portato all'eliminazione di tale dimensione riguardante gli aspetti culturali legati alla condivisione dei valori, il senso di appartenenza, il sentirsi parte di una comunità riguarda il fatto che tali aspetti possono considerarsi condizioni tipiche delle società arcaiche (Durkheim 1893). Inoltre, sono stati esclusi tutti gli aspetti riguardanti l'ambito economico (disuguaglianze di

reddito, esclusione dal mercato del lavoro, etc.) in quanto considerati fattori esterni di influenza e non dimensioni della coesione sociale.

Al fine di “misurare” le dimensioni della coesione sociale individuate, si è provveduto a realizzare una selezione teoricamente orientata di 24 indicatori dall'indagine su scala europea denominata European Social Survey (ESS, la rilevazione si riferisce alla sesta ondata). In tab.3.2 si riportano gli indicatori utilizzati per rilevare ciascuna delle sette dimensioni della coesione sociale suddivise in base all'ambito preso in considerazione e alla prospettiva utilizzata.

Tab.3.2 – Indicatori della coesione sociale

Indicatori	Prospettiva Soggettiva	Prospettiva Oggettiva
Micro	• Most people can be trusted or you can't be too careful • Most people try to take advantage of you, or try to be fair • Most of the time people helpful or mostly looking out for themselves • Feel people in local area help one another • Feel appreciated by people you are close to • Receive help and support from people you are close to	• How often socially meet with friends, relatives or colleagues • How many people with whom you can discuss intimate and personal matters • Take part in social activities compared to others of same age
Meso	• Immigration bad or good for country's economy • Country's cultural life undermined or enriched by immigrants • Immigrants make country worse or better place to live	• Worked in political party or action group last 12 months • Worked in another organisation or association last 12 months • Signed petition last 12 months • Involved in work for voluntary or charitable organisations, how often past 12 months
Macro	• Trust in country's parliament • Trust in the legal system • Trust in politicians • Trust in political parties	• State of health services in country nowadays • State of education in country nowadays • How satisfied with the national government • How satisfied with the way democracy works in country

Gli indicatori selezionati per ciascuna dimensione sono i seguenti:

- **Fiducia interpersonale:** all'intervistato veniva chiesto nel questionario usato nell' ESS, di rispondere ad una serie di domande riguardanti la fiducia riposta negli altri soggetti attraverso una scala (Cantril) il cui *range* variava da 0 (nessuna fiducia) a 10 (piena fiducia). Più precisamente, gli *items* risultano essere: “generally speaking, would you say that most people can be trusted, or that you can't be too careful in dealing with people?”, “Do you think that most people would try to take advantage of you if they got the chance, or would they try to be fair?”, “Would you say that most of the time people try to be helpful or that they are mostly looking out for themselves?”.
- **Supporto sociale:** l'intervistato doveva valutare in che misura si sentisse supportato ed apprezzato dalle persone a lui vicine e quanto ritenesse che le persone residenti nella sua zona si aiutassero mutualmente. Gli *items* sono: “to what extent do you feel that people in your local area help one another?”, “to what extent do you receive help and support from people you are close to when you need it?” (rilevati su una scala da 0 – per nulla – a 6 – completamente) e “to what extent do you feel appreciated by the people you are close to?” (rilevato tramite una scala da 0 – per nulla – a 10 – completamente).
- **Densità delle relazioni sociali:** I soggetti intervistati dovevano rispondere ad una serie di domande volte a rilevare quanto l'intervistato avesse contatti con altre persone, quante persone avesse con le quali discutere di questioni personali ed intime e in che misura

partecipasse ad attività sociali. Nella fattispecie le domande sono: “how often do you meet socially with friends, relatives or work colleagues?” (su una scala da 1 – mai – a 7 – ogni giorno), “how many people are there with whom you can discuss intimate and personal matters” (su una scala da 0 – nessuno – a 6 – dieci o più persone), e “compared to other people of your age, how often would you say you take part in social activities?” (su una scala da 1 – molto meno degli altri – a 5 – molto di più degli altri).

- Apertura: veniva chiesto alle persone di esprimere il loro giudizio sui soggetti immigrati attraverso domande che vertevano sui benefici o gli svantaggi che l'immigrazione può generare. Nello specifico gli *items* utilizzati sono: “it is generally bad or good for country's economy that people come to live here from other countries?”, “country's cultural life is generally undermined or enriched by people coming to live here from other countries?” e “is country made a worse or a better place to live by people coming to live here from other countries?” tutti rilevati su una scala Cantril.
- Partecipazione: le persone dovevano rispondere ad una serie di domande afferenti alla sfera della partecipazione politica, sociale e civica. Le domande utilizzate sono: “have you worked in a political party or action group?”, “have you worked in another organisation or association?”, “have you signed a petition?” (dicotomiche 1 Sì, 2 No) e “how often did you get involved in work for voluntary or charitable organisations?” rilevata su una scala da 1 (mai) a 6 (almeno una volta al mese)²⁷.
- Fiducia nelle istituzioni: gli intervistati dovevano esprimere un giudizio relativo al grado di fiducia riposto in una serie di istituzioni (parlamento, sistema legale, uomini politici e partiti politici) aiutandosi con una scala Cantril (0 = nessuna fiducia, 10 = totale fiducia).
Legittimità delle istituzioni: i soggetti dovevano esprimere un giudizio relativo al grado di soddisfazione nei confronti dell'operato del governo e valutare lo stato di una serie di servizi pubblici. Nello specifico gli *items* sono: “thinking about the country's government, how satisfied are you with the way it is doing its job?”, “how satisfied are you with the way democracy works in your country?” su una scala da 0 (estremamente insoddisfatto) a 10 (estremamente soddisfatto) e “please say what you think overall about the state of education in your country nowadays”, “please say what you think overall about the state of health services in your country nowadays” su una scala da 0 (pessimo) a 10 (eccellente).

In riferimento a quest'ultima dimensione la prospettiva viene definita oggettiva in senso lato perché si riferisce soprattutto a comportamenti ed esperienze che nel loro insieme possono essere considerate *proxies* di un punto di vista oggettivo, anche se riferite ai soggetti.

3.1.3 Dati, obiettivi e procedure

La base di dati utilizzata per effettuare le nostre analisi riguarda l'indagine denominata European Social Survey. L'ESS è un'indagine svolta all'interno degli Stati europei che viene condotta ogni due anni a partire dal 2002. Ha come obiettivo generale la rilevazione degli atteggiamenti, credenze e comportamenti dei cittadini europei. Il *dataset* da noi utilizzato riguarda la sesta ondata (2012). In particolare, i dati sono stati raccolti all'interno di 29 Paesi. In totale l'ampiezza campionaria è di 54672

²⁷ L'*item* originale era rilevato su una scala da 1 (almeno una volta al mese) a 6 (mai), quindi si è provveduto ad invertirne la polarità.

intervistati. Il campione risulta essere rappresentativo per gli individui da 15 anni in su (senza limite massimo) residenti in un nucleo familiare all'interno di ogni Stato²⁸. Il campionamento per quote non viene consentito a nessuno stadio della procedura che risulta essere puramente casuale; inoltre, non è prevista nessuna forma di sostituzione nel caso di mancata intervista (sia essa causata da un rifiuto, ineleggibilità o un non contatto dovuto a varie motivazioni). L'età media del campione risulta essere di 48,3 anni (*Dev.Standard* = 18,5) e il 54,4% è costituito da donne.

Gli obiettivi della nostra indagine sono molteplici. Il primo consiste nel testare empiricamente la validità della definizione operativa proposta. In secondo luogo, si tenterà di fornire una definizione operativa multilivello della coesione sociale con l'obiettivo di verificare se il costrutto svolge le stesse funzioni a livello individuale e aggregato, testandone, quindi, l'isomorfismo ai due livelli. Un ulteriore obiettivo è individuabile nel tentativo di testare l'ipotesi che il modello di coesione sociale proposto sia invariante all'interno dei 29 Stati analizzati. Ciò equivale a dire che i meccanismi che generano coesione sociale sono i medesimi a prescindere dalla diversità dei contesti culturali presi in esame. Infine, si tenterà di specificare un modello ad equazioni strutturali multilivello nel tentativo di individuare i principali macro-fattori esterni che possono esercitare un'influenza sulla coesione sociale. In particolare, ci si soffermerà sugli aspetti legati alla diversità di etnie e alla sfera afferente l'ambito delle disuguaglianze all'interno di uno Stato.

Le tecniche di indagine a cui si farà ricorso al fine di rispondere agli interrogativi di indagine prefissati si riferiscono all'analisi fattoriale confermativa, all'analisi fattoriale confermativa multilivello, all'analisi fattoriale confermativa multilivello di secondo ordine, ai modelli ad equazioni strutturali multilivello e infine ai modelli di mediazione ad equazioni strutturali multilivello.

Tutte le analisi sono state condotte attraverso il ricorso ad Mplus 7. Tenendo in debito conto la diversa natura delle variabili (continue, ordinali e dicotomiche) le analisi sono state effettuate sulla matrice di covarianza utilizzando uno stimatore WLSMV – si veda il cap.2 – (Asparouhov e Muthén 2007; Muthén 1984). Il WLSMV è uno stimatore robusto che non assume che le variabili siano distribuite normalmente (distribuzione Gaussiana) ed è particolarmente adatto nel trattamento di variabili non continue. La bontà di adattamento dei modelli è stata valutata sia attraverso il classico ricorso alla statistica del chi-quadrato sia per mezzo di indici di bontà di adattamento alternativi come il *Comparative fit index* (CFI) e il *Root mean square error of approximation* (RMSEA) (Bentler 1992; Steiger e Lind 1980). Livelli accettabili di adattamento del modello ai dati sono considerati i valori di RMSEA inferiori a 0,06 e valori di CFI superiori a 0,9 (Hu and Bentler 1999; Hox 2010).

Nella tab.3.3 vengono sinteticamente riportate tutte le statistiche descrittive riferite agli indicatori della coesione sociale precedentemente esposti. Nell'appendice statistica si riportano tutte le statistiche complete relative alla distribuzione degli *items* utilizzati e alla numerosità campionaria suddivisa per i 29 Stati analizzati.

²⁸ Al link <http://www.europeansocialsurvey.org/methodology/sampling.html> sono disponibili le informazioni relative al piano di campionamento e alle procedure.

Tab.3.3 – Statistiche descrittive

	Tipo di variabile	Etichetta ^a	Media	D.S.	Min	Max	ICC	Slmm.	Curt.
Interpersonal trust									
Most people can be trusted or you can't be too careful	C	A3	4.92	2.489	0	10	.157	-.266	-.663
Most people try to take advantage of you, or try to be fair	C	A4	5.50	2.342	0	10	.139	-.406	-.329
Most of the time people helpful or mostly looking out for themselves	C	A5	4.86	2.385	0	10	.130	-.183	-.556
Density of social relations									
How often socially meet with friends, relatives or colleagues	O	C2	4.79	1.634	1	7	.102	-.428	-.771
How many people with whom you can discuss intimate and personal matters	O	C3	2.67	1.440	0	6	.119	.151	-.514
Take part in social activities compared to others of same age	O	C4	2.67	.970	1	5	.042	.027	-.251
Social support									
Feel people in local area help one another	C	D21	3.65	1.583	0	6	.051	-.467	-.380
Feel appreciated by people you are close to	C	D29	7.82	1.802	0	10	.054	-1.015	1.300
Receive help and support from people you are close to	C	D36	4.96	1.253	0	6	.038	-1.511	2.418
Participation									
Worked in political party or action group last 12 months	D	B12	.04	.188	0	1	.051	4.918	22.186
Worked in another organisation or association last 12 months	D	B13	.13	.338	0	1	.254	2.186	2.778
Signed petition last 12 months	D	B15	.19	.394	0	1	.215	1.564	.446
Involved in work for voluntary or charitable organisations, how often past 12 months	O	D1	1.96	1.581	1	6	.090	.584	-1.660
Openness									
Immigration bad or good for country's economy	C	B32	4.87	2.555	0	10	.073	-.112	-.551
Country's cultural life undermined or enriched by immigrants	C	B33	5.51	2.627	0	10	.127	-.301	-.556
Immigrants make country worse or better place to live	C	B34	4.94	2.420	0	10	.111	-.109	-.303
Institutional trust									
Trust in country's parliament	C	B2	3.94	2.679	0	10	.212	.102	-.923
Trust in the legal system	C	B3	4.73	2.832	0	10	.266	-.130	-.994
Trust in politicians	C	B5	3.22	2.449	0	10	.214	.286	-.851
Trust in political parties	C	B6	3.22	2.426	0	10	.207	.293	-.792
Legitimacy of institutions									
How satisfied with the national government	C	B22	3.98	2.582	0	10	.188	.048	-.911
How satisfied with the way democracy works in country	C	B23	5.16	2.574	0	10	.233	-.321	-.669
State of education in country nowadays	C	B24	5.56	2.426	0	10	.187	-.406	-.454
State of health services in country nowadays	C	B25	5.16	2.649	0	10	.264	-.258	-.835

a = l'etichetta della variabile si riferisce alla corrispondente domanda all'interno del questionario dell'ESS Wave 6 consultabile pubblicamente sul sito dell'ESS; Simm = Simmetria; Curt = Curtosi; Tipo di variabile: C = continua, O = ordinale, D = dicotomica; ICC = coefficiente di correlazione intraclasse.

3.1.4 L'analisi fattoriale confermativa del modello di coesione sociale

Come numerosi ricercatori hanno suggerito (Muth  n 1994, Hox 2010), a causa della complessit   dei modelli di analisi fattoriale confermativa multilivello (MCFA),   consigliabile iniziare l'analisi di una struttura fattoriale sottostante ad un costrutto complesso partendo da modelli pi  semplici per poi incrementare progressivamente il livello di complessit  . Per tali ragioni, anche se da un punto di vista metodologico l'analisi non risulta pienamente corretta in quanto non viene presa in debito conto la natura gerarchica dei dati a disposizione (individui annidati gerarchicamente all'interno di Stati)²⁹, la prima analisi da noi effettuata si riferisce ad un'analisi fattoriale confermativa classica – sotto-modello dei pi  generali modelli ad equazioni strutturali dove le relazioni strutturali fra i costrutti non sono specificate. Tale analisi si pone l'obiettivo primario di testare se effettivamente sono individuabili empiricamente le sette dimensioni della coesione sociale elaborate teoricamente. L'approccio come detto   confermativo, ossia vi   l'ipotesi di partenza, in base alla quale la coesione sociale   composta da sette dimensioni fondamentali, che viene testata empiricamente. L'analisi fattoriale confermativa classica non tiene conto, come detto, della struttura multilivello dei dati. Ci  significa che le analisi

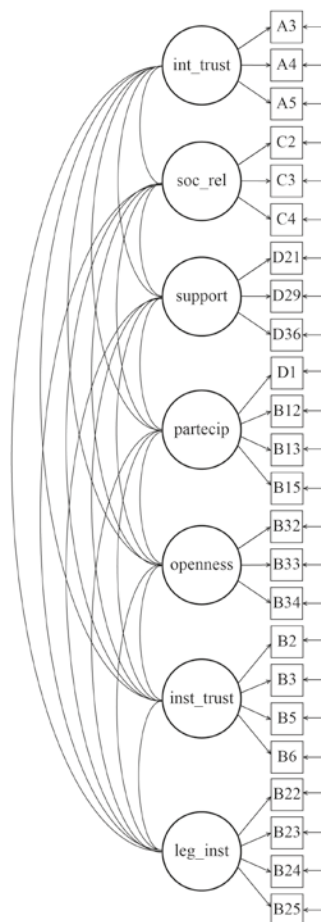
²⁹ Si veda il capitolo 2 in riferimento alle distorsioni che si generano quando si applicano tecniche che non tengono in considerazione la natura multilivello dei dati

vengono effettuate sulla matrice di covarianza totale e non vi è nessuna scomposizione della matrice in una porzione *between* ed in una *pooled within* (cfr. cap.2).

Il metodo di stima utilizzato è il cosiddetto *Weighted Least Squares Means and Variance Adjusted* (WLSMV).

Il modello di analisi fattoriale confermativa ad un livello è così composto: sette fattori latenti continui – ognuno riferito ad una delle sette dimensioni della coesione sociale – e 24 variabili dipendenti – gli indicatori prima illustrati (fig.3.1). Basandoci esclusivamente sulla statistica del chi-quadrato, che per questo modello risulta essere $\chi^2 = 9613.63$ (df 225, $p < .001$), il modello dovrebbe essere rifiutato. Ma numerose ricerche (Joreskog e Sorbom 1993; MacCallum et al 1996) hanno evidenziato l'estrema sensibilità del chi-quadrato all'ampiezza del campione. Più il campione è ampio e più il chi-quadrato tende ad aumentare. Un campione di 1000/1500 unità in letteratura viene già considerato molto grande, tenendo conto di tale aspetto e del nostro campione che risulta essere davvero enorme (54672 unità) tale valore del chi-quadrato è pienamente atteso. In questo caso, quindi, la statistica del chi-quadrato non è una misura attendibile della bontà di adattamento del nostro modello.

Fig.3.1 – Modello di analisi fattoriale della coesione sociale ad un livello



Per tali ragioni, sono stati considerati altri indici di adattamento proposti in letteratura che hanno la caratteristica di non essere influenzati dall'ampiezza del campione.

Gli indici alternativi mostrano un buon adattamento del modello. L'RMSEA di .028 (90% C.I. .027-.028) è considerevolmente al di sotto del valore soglia di .06 suggerito da Hu e Bentler (1999).

Allo stesso modo, il *Comparative fit index* (CFI) di .96 è maggiore del valore soglia di .95 che rappresenta un eccellente adattamento (Hu e Bentler 1999). In aggiunta, tutti i *factor loadings* sono significativamente diversi da zero ($p < .001$). I pesi fattoriali standardizzati per il costrutto *fiducia interpersonale* variano fra .701 e .771, da .415 a .651 per il fattore *densità delle relazioni sociali*, da .428 a .559 per il *supporto sociale*, da .424 a .885 per la dimensione *partecipazione*, da .776 a .853 per il costrutto *apertura*, da .795 a .866 per il fattore *fiducia nelle istituzioni* e, infine, da .569 a .815 per la dimensione *legittimità delle istituzioni*.

Possiamo concludere che il modello teorico della coesione sociale a sette dimensioni risulta essere sostenibile alla luce dell'adattamento del modello ai dati e che gli indicatori scelti risultano essere indicatori validi delle rispettive dimensioni che si intendevano misurare.

3.1.5 Analisi fattoriale confermativa Multilivello

In presenza di una struttura di dati annidati gerarchicamente (nel nostro caso individui all'interno di Paesi), ignorare la struttura multilivello può condurre a numerose distorsioni inficiando i risultati della ricerca (Hox, 2010; Snijder e Bosker 1999). Con gli individui annidati in unità di secondo livello le osservazioni risultano essere non più indipendenti (l'indipendenza delle osservazioni è uno degli assunti primari di tutti i modelli lineari generalizzati). In tal modo, gli errori standard vengono sottostimati e i test statistici di significatività risultano non più attendibili, ovvero gli errori del Tipo 1 ricorrono con più frequenza (Raudenbush e Bryk 2002). I modelli multilivello in tali situazioni si configurano come l'approccio corretto al fine di analizzare strutture di dati *nested*. Inoltre, l'analisi multilivello ci offre l'opportunità teorica di distinguere ed individuare un effetto individuale e un effetto di gruppo (o effetto di secondo livello).

I modelli multilivello ad equazioni strutturali (MLSEM) sono una generalizzazione dei classici modelli ad equazioni strutturali (SEM). Una delle principali sottoclassi dei MLSEM sono i modelli di analisi fattoriale confermativa (MCFA) che consentono la stima di modelli di misurazione notevolmente complessi (Mehta e Neale 2005). Essi permettono di definire simultaneamente un modello di misurazione individuale attraverso il ricorso ad indicatori osservati o costrutti latenti di primo livello e un modello di misurazione di gruppo o di secondo livello (*cluster-level measurement model*) per mezzo di indicatori osservati o costrutti latenti di secondo livello e consentono alle intercette degli indicatori di primo livello di assumere la forma di variabili casuali, ossia di essere *random* (Muthén 1994; Hox 2010).

La MCFA suddivide la matrice di covarianza totale in una matrice di covarianza fra i gruppi e una entro i gruppi (per approfondimenti si rimanda al capitolo 2). Queste due matrici di covarianza rappresentano la base di dati su cui si effettuano le analisi al fine di individuare una struttura fattoriale per ogni singolo livello (Muthén 1994; Hox 2010; Cheung e Au 2005).

Per chiarezza, riportiamo nuovamente il modello di analisi fattoriale a due livelli che assume la seguente forma (si veda anche Muthén 1991; Muthén 1994; Mehta e Neale 2005; Hox 2013; Selig et al. 2008):

$$y_{ij} = \mu_j + \lambda_w \eta_{wij} + \varepsilon_{wij}$$

$$\mu_j = \mu + \lambda_b \eta_{bj} + \varepsilon_{bj}$$

La porzione individuale del modello è rappresentata dalla prima equazione, mentre la seconda equazione riproduce il modello di secondo livello. Le due equazioni sono connesse dal termine μ_j che rappresenta l'intercetta di una unità di secondo livello J , nel nostro caso l'intercetta di uno dei 29 Paesi presenti nell'analisi. Tale intercetta rappresenta simultaneamente anche una variabile dipendente all'interno dell'equazione di secondo livello. Pertanto, le intercette degli indicatori di ogni Stato si configurano tutte come variabili *random*, ossia variabili a cui è permesso variare all'interno delle unità di secondo livello. Il fattore o costrutto latente di secondo livello, rappresentato dal termine η_{bj} , dà conto della varianza delle intercette di secondo livello; la varianza di secondo livello residuale – ossia quella che rimane “non spiegata” dopo aver controllato per l'effetto del fattore di secondo livello – è contenuta nel termine di errore di secondo livello, ossia ϵ_{bj} .

Prima di specificare il nostro modello di analisi fattoriale confermativa multilivello della coesione sociale, abbiamo esaminato la variabilità di primo livello (individuale) e di secondo livello (Stato) di ogni indicatore. Il coefficiente di correlazione intraclasse (ICC) fornisce al contempo sia una stima della variabilità di secondo livello di un indicatore che il grado di non-indipendenza delle osservazioni individuali annidate all'interno dei Paesi. Più precisamente l'ICC indica la variazione delle medie di ogni indicatore fra le unità di secondo livello suddivisa per la varianza totale (l'ICC varia da 0 – nessuna varianza di secondo livello – a 1 – la varianza è tutta di secondo livello, ossia non vi è varianza di primo livello). Non vi è un limite prestabilito e nemmeno accordo su quanto il valore dell'ICC debba essere grande per essere accettato come sufficiente, tuttavia la maggioranza dei ricercatori concordano che l'ICC dovrebbe essere almeno superiore a .05 (Hox 2010; Snijder e Bosker 1999).

L'ICC per ogni indicatore da noi selezionato è illustrato nell'ultima colonna della tab.3.3 dove si riportano le statistiche descrittive. Con l'eccezione di soli due indicatori su 24 (“Take part in social activities” e “Receive help and support”), gli *items* mostrano un valore dell'ICC maggiore del valore soglia consigliato in letteratura assumendo valori che variano da .266 (ossia più di un quarto della varianza è dovuto al Paese di appartenenza) fino a .05. L'ICC medio risulta quantificabile in .146, ossia il 15% della variabilità è dovuta alle unità di secondo livello. Tali valori indicano che vi è sufficiente variabilità di secondo livello tale da giustificare e rendere necessario l'utilizzo di un approccio multilivello. Se non si tiene conto di tale struttura dei dati non solo, come detto, si generano distorsioni inficiando l'efficienza di tutte le stime, ma si perde anche una considerevole opportunità teorica rinunciando a spiegare (o riprodurre) la porzione di varianza di secondo livello³⁰.

Dopo aver mostrato la logica sottostante ai modelli di analisi fattoriale confermativa di secondo livello ed aver illustrato le motivazioni metodologiche che hanno condotto alla scelta di tale tecnica passeremo a presentare il modello della coesione sociale multilivello.

Sono stati stimati tre modelli, ognuno volto a testare una precisa ipotesi, che per semplicità chiameremo Modello-1, Modello-2 e Modello-3.

³⁰ Tecnicamente la varianza di secondo livello può ridursi anche attraverso l'introduzione di variabili di primo livello.

3.1.6 Modello 1: CFA multilivello

Nel primo modello di analisi fattoriale confermativa multilivello si è cercato innanzitutto di verificare se la struttura latente a sette fattori della coesione sociale identificata nel modello di CFA ad un livello fosse ancora presente. Pertanto è stato specificato un modello dalla seguente struttura: sette fattori latenti continui di primo livello (le sette dimensioni della coesione sociale), 24 variabili dipendenti di primo livello (gli indicatori osservati), sette fattori latenti continui di secondo livello (le sette dimensioni della coesione sociale riprodotte come costrutti di secondo livello) e 24 variabili latenti³¹ *random* di secondo livello (fig.3.2). Lo stimatore è nuovamente il WLSMV.

Il valore dell'RMSEA pari a .028 e quello dello CFI di .97 rivelano un eccellente adattamento del modello (tab.3.4). Lo *Standardized Root Mean Square Residual* (SRMR) indica inoltre un miglior adattamento del modello di primo livello rispetto al modello di secondo livello (SRMR within = .038, SRMR between = .075). Tutti i *factor loadings* di primo livello sono statisticamente significativi per $p < .001$. Anche i pesi fattoriali di secondo livello sono tutti significativi eccetto per gli *items* etichettati con D21, B12, B13 (tab.3.5). La non significatività di soli tre *factor loadings* di secondo livello in un modello così specificato costituiscono un risultato atteso dal momento che l'ampiezza campionaria di secondo livello (29 Paesi) non è molto elevata.

Tab.3.4 – Bontà di adattamento dei tre modelli di coesione sociale

	Modello-1: sette fattori al primo livello e sette fattori di secondo livello. Pesi fattoriali liberamente stimati	Modello-2: sette fattori al primo livello e sette fattori di secondo livello. Pesi fattoriali costretti ad essere uguali al primo e secondo livello	Modello-3: sette fattori al primo livello e sette fattori di secondo livello. Pesi fattoriali costretti ad essere uguali e varianze residue di secondo livello fissate a zero
χ^2	19853.762	18078.354	18034.812
df	462	486	503
CFI	.970	.973	.973
RMSEA	.028	.026	.025
SRMR Within	.038	.039	.039
SRMR Between	.075	.235	.284

I pesi fattoriali standardizzati di primo livello (tab.3.5) per il costrutto *fiducia interpersonale* variano da .666 a .739, per il fattore latente *densità delle relazioni sociali* da .442 a .573, per la dimensione *supporto sociale* variano da .483 a .629, da .434 a .878 per la *partecipazione*, da .783 a .844 per la dimensione denominata *apertura*, per il costrutto *fiducia nelle istituzioni* da .705 a .859 e, infine, i *factor loadings* della dimensione *legittimità delle istituzioni* sono contenuti fra .531 e .771.

Tab.3.5 – Stime dei parametri non standardizzati e standardizzati per il Modello-1

Indicatori	Pesi fattoriali non standardizzati (S.E.)		Pesi fattoriali standardizzati (S.E.)	
	Livello-1	Livello-2	Livello-1	Livello-2
<i>Fiducia interpersonale</i>				
A3	1.00 (-----)	1.00 (-----)	0.73 (0.001)	0.94 (0.033)
A4	0.89 (0.002)	0.96 (0.109)	0.69 (0.001)	1.02 (0.025)
A5	0.87 (0.002)	0.85 (0.127)	0.66 (0.001)	0.92 (0.045)
<i>Densità delle relazioni sociali</i>				
C2	1.00 (-----)	1.00 (-----)	0.44 (0.004)	0.57 (0.111)
C3	1.13 (0.019)	1.64 (0.694)	0.57 (0.005)	0.99 (0.115)

³¹ Tali variabili vengono a rigore definite latenti in quanto non sono direttamente osservate.

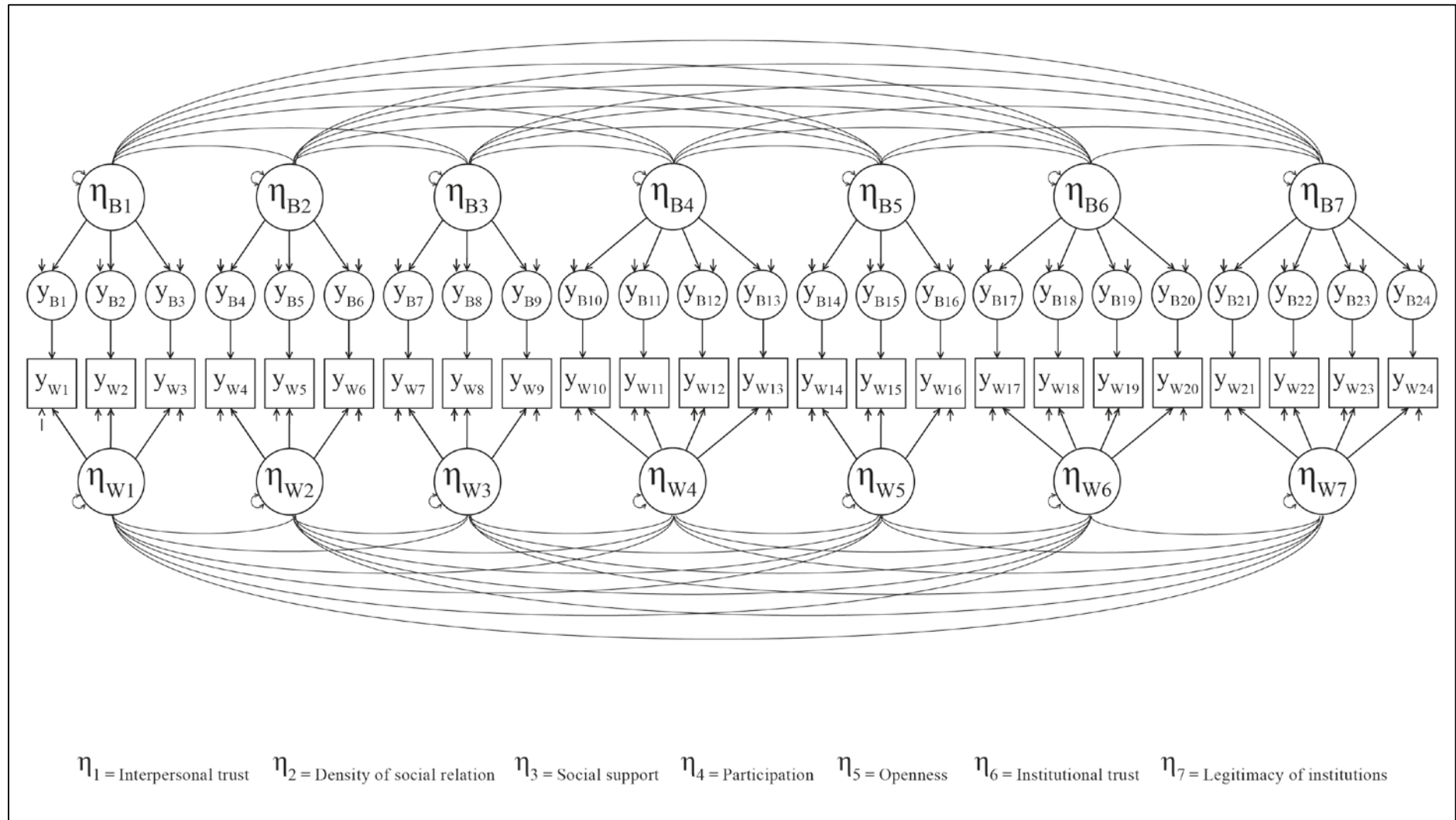
C4	0.94 (0.024)	0.57 (0.268)	0.54 (0.009)	0.69 (0.211)
<i>Supporto sociale</i>				
D29	1.00 (-----)	1.00 (-----)	0.48 (0.003)	0.38 (0.143)
D21	1.14 (0.013)	1.50 (0.840)	0.62 (0.003)	0.68 (0.129)
D36	0.71 (0.007)	1.23 (0.601)	0.49 (0.003)	0.82 (0.095)
<i>Partecipazione</i>				
D1	1.00 (-----)	1.00 (-----)	0.43 (0.010)	0.87 (0.227)
B12	1.18 (0.071)	0.38 (0.221)	0.61 (0.015)	0.54 (0.127)
B13	2.80 (0.544)	2.76 (1.340)	0.87 (0.042)	0.94 (0.297)
B15	1.26 (0.098)	1.48 (0.829)	0.63 (0.033)	0.90 (0.201)
<i>Apertura</i>				
B32	1.00 (-----)	1.00 (-----)	0.78 (0.001)	0.95 (0.114)
B33	0.99 (0.001)	1.32 (0.422)	0.78 (0.001)	0.93 (0.083)
B34	0.99 (0.001)	1.12 (0.337)	0.84 (0.001)	0.91 (0.074)
<i>Fiducia nelle istituzioni</i>				
B2	1.00 (-----)	1.00 (-----)	0.84 (0.000)	0.99 (0.019)
B3	0.85 (0.001)	1.15 (0.149)	0.70 (0.001)	0.96 (0.030)
B5	0.93 (0.001)	0.89 (0.165)	0.85 (0.000)	0.96 (0.037)
B6	0.88 (0.001)	0.86 (0.148)	0.82 (0.000)	0.95 (0.042)
<i>Legittimità delle istituzioni</i>				
B22	1.00 (-----)	1.00 (-----)	0.77 (0.000)	0.82 (0.094)
B23	0.95 (0.002)	1.34 (0.283)	0.76 (0.001)	1.002 (0.035)
B24	0.64 (0.002)	0.95 (0.318)	0.53 (0.001)	0.84 (0.071)
B25	0.68 (0.002)	1.24 (0.511)	0.54 (0.001)	0.85 (0.137)

Nota: il primo indicatore di ciascun fattore è fissato ad 1 al fine di assegnare l'unità di misura al fattore latente

I *factor loadings* standardizzati di secondo livello (tab.3.5) variano fra .928 to 1.022 per il costrutto *fiducia interpersonale*, per la dimensione *densità delle relazioni sociali* da .577 a .995, per la dimensione *supporto sociale* variano da .389 a .821, da .547 a .940 per la *partecipazione*, da .916 a .950 per il costrutto *apertura*, per il fattore *fiducia nelle istituzioni* da .950 a .992 e, infine, i *factor loadings* della dimensione *legittimità delle istituzioni* sono contenuti fra .828 e 1.002. Da come possiamo notare, anche se di natura infinitesimale, due pesi fattoriali standardizzati di secondo livello sono maggiori dell'unità. Questo può accadere, ed in genere accade, quando le varianze residue di secondo livello vengono stimate assumendo un valore negativo. Come Hox afferma (2010), molte volte è necessario fissare a zero le varianze residue di secondo livello quando la numerosità campionaria di secondo livello è poco ampia e la varianza vera (ossia depurata dal termine di errore) di secondo livello è vicina allo zero. Inoltre, un coefficiente standardizzato più grande dell'unità, come fa notare Joreskog (1999), non necessariamente suggerisce che si è prodotta qualche distorsione nell'analisi³².

³² Si dovrebbe sempre fissare a zero le varianze residue di secondo livello vicine allo zero, ma nel nostro primo modello abbiamo preferito stimarle liberamente per valutarne l'ampiezza e la rilevanza delle stime fornite. Nel prossimo modello, il modello-2, tali varianze residue saranno fissate a zero.

Fig.3.2 – Modello multilivello della coesione sociale



3.1.7 Modello 2: Ipotesi di invarianza cross-level con il vincolo dell'eguaglianza dei factor loadings

Dal momento che il Modello-1 si configura come modello contraddistinto da due livelli di analisi (individui e Paesi), si solleva la questione fondamentale se la coesione sociale sia misurata allo stesso modo a livello individuale e a livello di Stati (Selig et al. 2008; Jak et al. 2013). Tale quesito chiama in causa il concetto di *cross-level invariance*, ossia l'invarianza del modello di misurazione della coesione sociale fra i due livelli di analisi. Al fine di testare l'ipotesi di *cross-level invariance* è stato specificato un ulteriore modello. Se tale modello – che ha la caratteristica di avere i *factor loadings* di primo e secondo livello condizionati ad assumere la stessa stima – presenta un buon adattamento ai dati è ragionevole concludere che i costrutti sono misurati analogamente a livello individuale e a livello di Paese. In aggiunta a tale aspetto fondamentale, stabilire l'uguaglianza dei pesi fattoriali fra i livelli di analisi ci permette di comparare direttamente le varianze latenti appartenenti ai due livelli individuando quanta porzione di varianza nei fattori latenti sia dovuta al livello individuale e quanta invece al livello dei Paesi. Per testare tale ipotesi si è proceduto a stimare il Modello-2 che risulta essere esattamente uguale al Modello-1 ma in aggiunta contiene una serie di costrizioni imposte sulle matrici di covarianza: ossia i pesi fattoriali di primo e di secondo livello sono costretti ad assumere stime perfettamente uguali. Lo stimatore rimane sempre il WLSMV.

Il Modello-2 mostra un ottimo adattamento ai dati con un valore di RMSEA di .026 e del CFI di .97 (tab.3.4). L'indice SRMR indica nuovamente un miglior adattamento del modello di primo livello rispetto al modello di secondo livello (tab.3.4). Le stime dei pesi fattoriali riferiti al Modello-2 risultano essere tutte significative per $p < .001$ sia a livello-1 che a livello-2 e i *factor loadings* standardizzati variano da .44 a .88 per il livello individuale e da .46 a .96 per il livello aggregato (si veda tab.3.6 per le stime complete standardizzate e non standardizzate dei pesi fattoriali)

Tab.3.6 – Stime non standardizzate e standardizzate dei pesi fattoriali riferiti al Modello-2 (pesi fattoriali fissati come uguali al primo e al secondo livello)

Indicatori	Pesi fattoriali non standardizzati (S.E.)	Pesi fattoriali standardizzati (S.E.)	
	Livello-1 e Livello-2	Livello-1	Level-2
<i>Fiducia interpersonale</i>			
A3	1.00 (-----)	0.73 (0.001)	0.95 (0.062)
A4	0.89 (0.002)	0.69 (0.001)	0.96 (0.059)
A5	0.87 (0.002)	0.66 (0.001)	0.96 (0.071)
<i>Densità delle relazioni sociali</i>			
C2	1.00 (-----)	0.44 (0.004)	0.54 (0.101)
C3	1.13 (0.019)	0.57 (0.005)	0.64 (0.147)
C4	0.93 (0.025)	0.54 (0.010)	1.00 (-----)
<i>Supporto sociale</i>			
D29	1.00 (-----)	0.48 (0.003)	0.46 (0.120)
D21	1.14 (0.012)	0.62 (0.003)	0.62 (0.141)
D36	0.71 (0.007)	0.49 (0.003)	0.57 (0.094)
<i>Partecipazione</i>			
D1	1.00 (-----)	0.44 (0.009)	0.55 (0.124)
B12	1.11 (0.062)	0.59 (0.014)	1.00 (-----)
B13	2.84 (0.544)	0.88 (0.038)	0.59 (0.130)
B15	1.28 (0.110)	0.65 (0.035)	0.49 (0.064)
<i>Apertura</i>			
B32	1.00 (-----)	0.78 (0.001)	1.00 (-----)
B33	0.99 (0.001)	0.78 (0.001)	0.77 (0.073)
B34	0.99 (0.001)	0.84 (0.001)	0.89 (0.084)
<i>Fiducia nelle istituzioni</i>			

B2	1.00 (-----)	0.84 (0.000)	1.00 (-----)
B3	0.85 (0.001)	0.70 (0.001)	0.74 (0.090)
B5	0.93 (0.001)	0.85 (0.000)	1.00 (-----)
B6	0.88 (0.001)	0.82 (0.000)	1.00 (-----)
<i>Legittimità delle istituzioni</i>			
B22	1.00 (-----)	0.77 (0.000)	1.00 (-----)
B23	0.95 (0.002)	0.76 (0.001)	0.92 (0.092)
B24	0.64 (0.002)	0.53 (0.001)	0.74 (0.114)
B25	0.68 (0.002)	0.54 (0.001)	0.60 (0.118)

Nota: il primo indicatore di ciascun fattore è fissato ad 1 al fine di assegnare l'unità di misura al fattore latente; gli altri pesi fattoriali a livello *between* nella soluzione standardizzata assumono valore 1 perché le rispettive varianze residue sono fissate a zero.

Dal momento che le statistiche utili a valutare la bontà di adattamento del modello indicano che il Modello-2 si adatta ai dati in modo ottimale (mostrando indici migliori del Modello-1), l'ipotesi prima descritta dell'uguaglianza dei pesi fattoriali fra i due livelli di analisi, e quindi l'uniformità del modello di misurazione della coesione sociale al primo e al secondo livello, è ragionevolmente sostenibile. La condizione di invarianza del modello di misurazione implica che i fattori di primo livello e quelli di secondo livello hanno la stessa interpretazione. In altre parole, gli *items* caricano allo stesso modo e con la stessa intensità su ciascuno dei rispettivi costrutti di primo e secondo livello. Data l'invarianza dei pesi fattoriali – aspetto che determina che l'unità di misura riferita ai fattori di primo e di secondo livello è la stessa – possiamo comparare direttamente le varianze dei sette costrutti latenti (Mehta e Neale 2005). A tale scopo, abbiamo calcolato gli ICC delle sette dimensioni latenti della coesione sociale (tab.3.7): il 24% della variabilità nel costrutto *fiducia interpersonale* è dovuta al Paese al quale gli intervistati appartengono, il che equivale a dire che il 24% del fenomeno è di tipo aggregato, ossia un effetto macro. Allo stesso modo quasi il 15% della variabilità nel fattore *densità delle relazioni sociali* è dovuto al livello macro, il 5% nella dimensione *supporto sociale*, il 13,5% nel fattore *partecipazione*, il 13% in *apertura*, quasi il 29% nel costrutto *fiducia nelle istituzioni* e, infine, ben il 31% nella dimensione *legittimità delle istituzioni*.

Tenendo presente questi dati, si può affermare con una certa ragionevolezza che le unità macro di secondo livello, ossia i Paesi, assumono un ruolo fondamentale nel processo di influenza degli atteggiamenti e dei comportamenti degli individui relativamente ai meccanismi di generazione della coesione sociale.

Tab.3.7 – ICC dei costrutti latenti

Costrutti	ICC
<i>Fiducia interpersonale</i>	.239
<i>Densità delle relazioni sociali</i>	.146
<i>Supporto sociale</i>	.052
<i>Partecipazione</i>	.135
<i>Apertura</i>	.130
<i>Fiducia nelle istituzioni</i>	.286
<i>Legittimità delle istituzioni</i>	.309

3.1.8 Modello 3: Ipotesi di invarianza di misurazione del modello di coesione sociale

L'invarianza del modello di misurazione è un aspetto di primissimo piano all'interno del contesto delle ricerche comparative, *cross-cultural* o *cross-national*, come l'indagine che qui stiamo esponendo. Data l'elevata eterogeneità dei contesti in cui tali indagini vengono effettuate possono sollevarsi numerosi problemi relativamente a molteplici aspetti come problemi di traduzione del questionario, *bias* culturali, differenti interpretazioni delle domande del questionario che possono generare numerose distorsioni. In aggiunta l'invarianza di misurazione fra i diversi gruppi, culture, contesti, Stati si configura come condizione necessaria al fine della comparazione di tali unità sulle dimensioni latenti individuate. Se la struttura fattoriale e il modello di misura delle variabili latenti varia al variare del contesto non è più possibile effettuare nessuna forma di comparazione in quanto le differenze nei coefficienti strutturali possono essere semplicemente il riflesso di un modello di misura diverso piuttosto che riferirsi ad una differenza sostantiva. A tal proposito, è possibile individuare tre livelli progressivi di invarianza: configurativa, metrica e scalare (Meredith 1993; Cieciuch et al. 2104; Davidov et al. 2014). Il livello più basso di invarianza – l'invarianza configurativa – sussiste nel momento in cui gli stessi *items* caricano sugli stessi costrutti latenti in tutte le unità di secondo livello (ad esempio, Stati, regioni, classi scolastiche), ossia abbiamo un'identica struttura fattoriale all'interno di tutti i gruppi. L'invarianza di configurazione non garantisce però in nessun modo che tali *items* siano in aggiunta misurati allo stesso modo, qualcuno potrebbe risultare essere un indicatore valido in un gruppo e meno valido in un altro gruppo. Pertanto, testare l'ipotesi di invarianza metrica è una condizione necessaria se si vogliono comparare i coefficienti di regressione e le covarianze fra i diversi gruppi. L'invarianza scalare, invece, richiede il soddisfacimento di un ulteriore requisito ossia che le intercette di ogni singolo *items* siano uguali fra tutte le unità di secondo livello. Questa condizione – l'invarianza delle intercette in aggiunta alle precedenti – si pone come necessaria se si vogliono comparare le medie dei fattori latenti fra i diversi gruppi di secondo livello.

Diversamente dalle tecniche classiche utilizzate per testare le ipotesi relative all'invarianza di misurazione, sia essa configurativa, metrica o scalare, (per esempio la tecnica dell'analisi multi-gruppo attraverso il ricorso ad un modello ad equazioni strutturali), in un modello di analisi fattoriale confermativa multilivello non è possibile discernere fra invarianza metrica e scalare (Jak et al. 2013). In tali modelli, il test per valutare l'ipotesi di invarianza di misurazione è effettuato attraverso l'aggiunta di alcune restrizioni alle matrici di covarianza: i parametri riferiti ai pesi fattoriali vengono condizionati ad essere uguali fra i due livelli di analisi e le varianze residue di secondo livello vengono fissate a zero.

Con lo scopo di continuare nella valutazione dell'invarianza del modello di misurazione della coesione sociale, abbiamo specificato un ulteriore modello (Modello-3) aggiungendo alle matrici di covarianza ulteriori limitazioni riferibili alle varianze residue di secondo livello fissate a zero. Lo stimatore rimane il WLSMV. Gli indici di bontà di adattamento del modello indicano un buon adattamento ai dati, con l'RMSEA = .025 e il CFI = 0.97 (tab.3.4). L'indice SRMR rivela come il modello di primo livello mostri un miglior adattamento del modello di secondo livello (tab.3.4). Infine, tutti i *factor loadings* risultano essere significativi per $p < .001$ (tab.3.8).

Tab.3.8 – Stime non standardizzate e standardizzate dei pesi fattoriali riferiti al Modello-3 (pesi fattoriali fissati come uguali al primo e al secondo livello e varianze residue di secondo livello fissate a zero)

	Pesi fattoriali non standardizzati (S.E.)	Pesi fattoriali standardizzati (S.E.)
Indicatori	Livello-1 e Livello-2	Livello-1
<i>Fiducia interpersonale</i>		
A3	1.00 (-----)	0.73 (0.001)
A4	0.89 (0.002)	0.69 (0.001)
A5	0.87 (0.002)	0.66 (0.001)
<i>Densità delle relazioni sociali</i>		
C2	1.00 (-----)	0.44 (0.004)
C3	1.13 (0.019)	0.57 (0.005)
C4	0.93 (0.026)	0.53 (0.010)
<i>Supporto sociale</i>		
D29	1.00 (-----)	0.48 (0.003)
D21	1.14 (0.012)	0.62 (0.003)
D36	0.71 (0.007)	0.49 (0.003)
<i>Partecipazione</i>		
D1	1.00 (-----)	0.44 (0.010)
B12	1.09 (0.064)	0.59 (0.015)
B13	2.90 (0.578)	0.89 (0.039)
B15	1.30 (0.117)	0.65 (0.036)
<i>Apertura</i>		
B32	1.00 (-----)	0.78 (0.001)
B33	0.99 (0.001)	0.78 (0.001)
B34	0.99 (0.001)	0.84 (0.001)
<i>Fiducia nelle istituzioni</i>		
B2	1.00 (-----)	0.84 (0.000)
B3	0.85 (0.001)	0.70 (0.001)
B5	0.93 (0.001)	0.85 (0.000)
B6	0.88 (0.001)	0.82 (0.000)
<i>Legittimità delle istituzioni</i>		
B22	1.00 (-----)	0.77 (0.000)
B23	0.95 (0.002)	0.76 (0.001)
B24	0.64 (0.002)	0.53 (0.001)
B25	0.68 (0.002)	0.54 (0.001)

Note: dal momento che le varianze residue di secondo livello sono condizionate a zero, i coefficienti standardizzati di livello-2 sono tutti equiparati ad uno. Il primo indicatore di ciascun fattore è fissato ad 1 al fine di assegnare l'unità di misura al fattore latente

Tenendo presente che gli indici di adattamento non degradano passando dal Modello-2 al Modello-3 (si veda la tab.3.4) – il Modello-3 che è più restrittivo avendo maggiori gradi di libertà non mostra nessun deterioramento dell'adattamento generale – possiamo ragionevolmente sostenere che l'invarianza di misurazione del modello della coesione sociale qui proposto sembra supportata (o meglio, non rifiutata). L'invarianza di misurazione ci assicura che non vi sia nessun “cluster bias” (si veda a riguardo Jak et al. 2013). In altre parole, le relazioni che sussistono fra i fattori latenti e gli indicatori sono le stesse nei 29 Paesi oggetto di indagine – il modello di coesione sociale è uguale nei 29 Paesi considerati – e tale conformità sussiste anche passando dal primo livello al secondo livello. Più avanti approfondiremo tali elementi di assoluta rilevanza.

3.1.9 Analisi fattoriale confermativa multilivello di secondo ordine

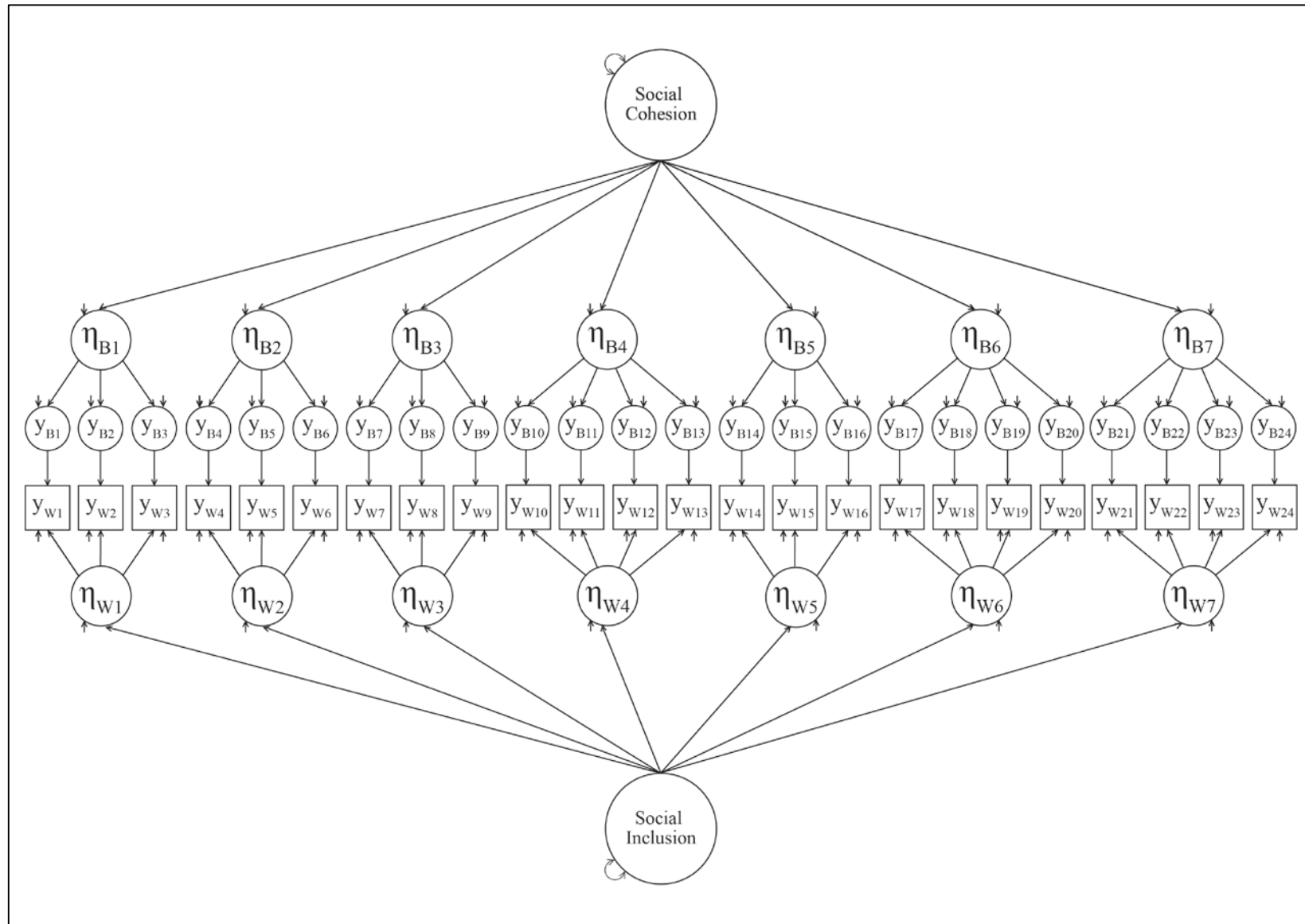
Le motivazioni che possono condurre, e giustificare, la specificazione di un modello di analisi fattoriale confermativa di secondo ordine sono molteplici. Tali ragioni hanno a che fare con il triplice aspetto riguardante la teoria, la statistica e la ricerca empirica. Da un punto di vista teorico risulta di assoluto interesse esaminare l'esistenza di un'eventuale struttura fattoriale di secondo ordine capace di sintetizzare i fattori di primo ordine individuati. Nel nostro caso, ciò si traduce nella possibilità ed eventualità di individuare un unico fattore di secondo ordine capace di dar conto delle variazioni dei sette costrutti di primo ordine individuati. L'individuazione di un fattore più generale in grado di spiegare l'intera struttura fattoriale di primo ordine si configura come un notevole passo in avanti in termini di semplicità ed eleganza del modello. Inoltre, l'eventuale presenza di un unico fattore di secondo ordine in grado di spiegare i sette costrutti individuati costituirebbe un indizio sostanziale della presenza di un costrutto ancor più generale, che potremmo per l'appunto chiamare *coesione sociale*, operante all'interno della società capace di dar conto dei fenomeni che osserviamo. Statisticamente, invece, è il *pattern* delle correlazioni fra i fattori latenti di primo ordine che suggerisce l'opportunità o meno di specificare un modello di secondo ordine. Infatti, nel caso in cui non sussista nessuna forma di relazione in termini di correlazioni, non vi sarebbe nessuna giustificazione accettabile nello stimare un modello di analisi fattoriale di ordine maggiore. Nonostante ciò che si è appena sostenuto, l'opportunità di specificare un modello che implichi la presenza di fattori di ordine maggiore deve essere una scelta esclusivamente guidata dalla teoria. Un modello di secondo ordine deve essere sempre teoricamente e concettualmente giustificato. Infatti, solo se la teoria suggerisce la presenza di un fattore di secondo ordine (o maggiore, ma è rarissimo) che può spiegare la covariazione fra i fattori di primo ordine, si può specificare un modello di analisi fattoriale confermativa di secondo ordine. Tale considerazione assume la giusta rilevanza se si tiene ben presente che i fattori di secondo ordine sono individuati e specificati senza il ricorso a nessun tipo di indicatore osservato.

Il nostro intero impianto teorico della coesione sociale, con la suddivisione del concetto in sette sotto-dimensioni costituenti, implica logicamente l'esistenza teorica di un fattore latente di ordine maggiore rispetto a quelli individuati. Per tale ragione, nel modello che sarà a breve presentato, ipotizziamo, trattandosi di un modello multilivello, l'esistenza di un unico fattore sia a livello individuale sia a livello aggregato responsabile delle covariazioni osservate fra i sette costrutti prima individuati.

Per testare tale ipotesi, è stato stimato un modello di analisi fattoriale confermativa multilivello di secondo ordine. Tale modello presenta un fattore di secondo ordine di primo livello, un fattore di secondo ordine di secondo livello, sette fattori latenti continui di primo ordine di primo livello (le sette sotto-dimensioni della coesione sociale), 23 variabili dipendenti di primo livello (gli indicatori osservati), sette fattori latenti continui di primo ordine di secondo livello (le sette dimensioni della coesione sociale riprodotte come costrutti di secondo livello) e 23 variabili latenti *random* di secondo livello³³ (fig.3.3).

³³ Al fine di far raggiungere un punto di convergenza del modello, è stata eliminata la variabile dicotomica "b13" in quanto mostrava errori standard estremamente grandi causati da problemi computazionali.

Fig.3.3 – Modello multilivello di secondo ordine della coesione sociale



La struttura fattoriale di primo ordine è stata lasciata invariata con i pesi fattoriali condizionati ad assumere la stessa stima fra i due livelli e le varianze residue di secondo livello sono state fissate a zero. In aggiunta a tali costrizioni, abbiamo ulteriormente condizionato i *factor loadings* di secondo ordine ad assumere gli stessi valori nel modello di primo livello e in quello di secondo livello. Lo stimatore a cui è stato fatto ricorso è nuovamente il WLSMV.

Cercando di tenere ben presente e in debita considerazione che il modello qui presentato è contraddistinto da estrema complessità e che si sta tentando di spiegare ben 42 covarianze fra i costrutti di primo ordine (sia di primo che di secondo livello) attraverso il ricorso ad un solo fattore di primo livello ed uno di secondo livello, gli indici di bontà di adattamento indicano tutti un ottimo adattamento del modello (RMSEA = .038 e CFI = 0.949). Bisogna considerare che un modello di analisi fattoriale di secondo ordine non può mai mostrare un adattamento ai dati migliore del modello di primo ordine in quanto un modello di secondo ordine ha lo scopo di riprodurre, attraverso il ricorso ad un numero più parsimonioso di parametri, le covarianze fra i fattori di primo ordine che, all'inverso, sono liberamente stimate nel modello di primo ordine. L'SRMR indica un miglior adattamento del modello di primo livello rispetto al modello di secondo livello (SRMR within = .056, SRMR between = .246³⁴).

I pesi fattoriali di primo e di secondo ordine di primo livello così come quelli di primo e secondo ordine di secondo livello sono tutti significativi per $p < .001$. Inoltre, ognuno dei fattori latenti di primo ordine carica in maniera abbastanza rilevante sui fattori di secondo ordine (tab.3.9). I pesi fattoriali standardizzati di secondo ordine di primo livello risultano essere rispettivamente: .533 per la *fiducia interpersonale*, .249 per il costrutto *densità delle relazioni sociali*, .378 per *supporto sociale*, .273 per la dimensione *partecipazione*, .408 per *apertura*, .807 per la dimensione *fiducia nelle istituzioni* e .874 per il fattore *legittimità delle istituzioni*.

Invece, i pesi fattoriali standardizzati di secondo ordine di secondo livello per ogni costrutto sono: .681 per la *fiducia interpersonale*, .412 per *densità delle relazioni sociali*, .955 per *supporto sociale*, .586 per *partecipazione*, .764 per *apertura*, .912 per *fiducia nelle istituzioni* e .909 per il fattore *legittimità delle istituzioni*.

Tab.3.9 – Pesi fattoriali di secondo ordine

Costrutti di primo ordine	Non standardizzati	Standardizzati	
	Livello-1 e Livello-2	Livello-1	Livello-2
<i>Fiducia interpersonale</i>	1.000	0.533	0.681
<i>Densità delle relazioni sociali</i>	0.184	0.249	0.412
<i>Supporto sociale</i>	0.348	0.378	0.955
<i>Partecipazione</i>	0.157	0.273	0.586
<i>Apertura</i>	0.874	0.408	0.764
<i>Fiducia nelle istituzioni</i>	1.792	0.807	0.912
<i>Legittimità delle istituzioni</i>	1.731	0.874	0.909

Dal momento che la bontà di adattamento nella soluzione fattoriale di secondo ordine (RMSEA = .038 e CFI = 0.949) appena presentata non decresce in maniera significativa comparata al modello di primo ordine, possiamo ragionevolmente concludere che il modello di analisi fattoriale di secondo ordine

³⁴ L'SRMR, diversamente rispetto ai SEM classici, non è un indice globale ma è calcolato separatamente per la parte *within* e per la parte *between*. Pertanto, non possiamo assumere che funzioni come nei SEM classici. Al momento vi sono poche indagini a riguardo. Pertanto, si è usato l'SRMR al solo scopo di comparare il *fit* del modello *within* e *between*.

multilivello dà conto in maniera soddisfacente delle covarianze rilevate fra i sette costrutti di primo ordine. Ciò costituisce un buon indizio della presenza di una proprietà più generica non osservata che possiamo appunto chiamare *coesione sociale* al livello di Paese e che a nostro parere corrisponde all' *inclusione sociale* a livello individuale (si veda par. successivo). Tali due costrutti danno conto dei fenomeni di superficie che osserviamo e sono responsabili di alcune delle dimensioni fondamentali dell'agire umano come la fiducia generale negli altri individui, la partecipazione, la disponibilità ad accettare la diversità e la densità delle relazioni sociali.

Calcolando l'ICC (che assume un valore di .347) riferito al costrutto generale *coesione sociale* notiamo come ben un terzo della variabilità del fenomeno in esame sia da attribuirsi al livello aggregato – ovvero il Paese in cui si risiede contribuisce per il 35% della varianza del costrutto *coesione sociale* – rivelando come un terzo del fenomeno sia spiegabile dalla società in cui si risiede. Il contesto macro esercita quindi un'influenza più che rilevante sugli atteggiamenti e sui comportamenti individuali, come meglio illustreremo nel prossimo paragrafo.

3.2 Le implicazioni del modello multilivello della coesione sociale

Negli ultimi venti anni si è assistito al rifiorire del concetto di coesione sociale che ha portato sia il mondo accademico che quello politico ad occuparsene con crescente interesse. Ciò ha determinato sicuramente il proliferare degli studi sulla coesione, così come visto nel capitolo 1, ma a dispetto dei numerosi tentativi di definire e concettualizzare la coesione sociale, gli studi empirici volti a fornire una definizione operativa rigorosa ed orientata empiricamente sono stati esigui (Dickes et al. 2010).

Numerosi ricercatori hanno considerato la coesione sociale un concetto complesso riferendosi ad esso come concetto multidimensionale (Jenson 1998; Bernard 1999; Berger-Schmitt 2000), ma si registrano pochi tentativi volti a mettere alla prova empirica tale ipotesi di multidimensionalità. In aggiunta, vi sono ancora meno studi empirici che affrontano il tema della coesione sociale nell'ambito di un *framework* teorico volto alla validazione del concetto in molteplici contesti culturali testando l'equivalenza del modello in numerosi Paesi. Inoltre, per quanto ne sappiamo, il nostro studio costituisce l'unica indagine volta a fornire una definizione operativa multilivello della coesione sociale attraverso la specificazione di un modello ad equazioni strutturali multilivello.

Nel nostro studio, conformandoci alla visione comune, la coesione sociale è stata considerata un concetto multidimensionale data la complessità che la contraddistingue, complessità che si è cercato di descrivere nel capitolo 1 dove sono state forniti i numerosi punti di vista sul tema, spesso in contrasto fra di loro.

La definizione operativa da noi proposta ha previsto la suddivisione del concetto generale in sette sotto-dimensioni costituenti. Le prime tre dimensioni (*fiducia interpersonale, densità delle relazioni sociali, supporto sociale*) si pongono come dimensioni volte a rappresentare la sfera afferente alle interazioni sociali che coinvolgono direttamente gli individui. Le dimensioni denominate *apertura* e *partecipazione* si riferiscono invece, rispettivamente, alla disponibilità ad accettare gruppi di soggetti provenienti da altre parti del mondo e all'impegno civico. Infine, le due dimensioni strettamente interconnesse riferite alla *legittimità delle istituzioni* e alla *fiducia nelle istituzioni* riguardano l'ambito riconducibile alla fiducia generalizzata nei confronti del buon funzionamento della società e delle istituzioni che la compongono.

Abbiamo visto nei paragrafi precedenti come siano stati presentati diversi³⁵ modelli della coesione sociale attraverso il ricorso ad un approccio multilivello ad equazioni strutturali che ci ha permesso sia di tener conto della struttura gerarchica dei dati sia dell'errore di misurazione, fornendoci in tal modo stime maggiormente affidabili.

Le motivazioni che ci hanno condotto al ricorso ad un approccio ad equazioni strutturali multilivello sono molteplici. Innanzitutto, il *dataset* utilizzato presentava una struttura di dati annidati gerarchicamente con gli individui contenuti in unità di secondo livello, pertanto le osservazioni non rispettavano l'assunto di indipendenza (l'ICC medio riferito agli *items* è risultato di .146, ossia quasi il 15% della varianza è di livello macro). In secondo luogo, si è tenuto conto del fatto che l'*European Social Survey* costituisce un'indagine di tipo *cross-national* che pone numerosi problemi di ordine metodologico. Un assunto critico presente in questo tipo di indagini comparative riguarda il fatto che lo strumento per la raccolta dei dati rilevi gli stessi costrutti alla stessa maniera in ogni singolo Paese preso in considerazione. Molto spesso tale condizione non si verifica in quanto possono prodursi numerose distorsioni soprattutto dovute a problemi di traduzione del questionario o a più generici *bias* culturali. L'approccio classico a cui si ricorre per testare l'invarianza di misurazione è la cosiddetta analisi fattoriale confermativa multi-gruppo (MGCFA). Tuttavia, tale approccio diventa impraticabile al punto da essere considerato quasi irrealizzabile quando il numero dei gruppi da prendere in considerazione è molto ampio, come nel nostro caso (Byrne e van De Vijver 2010). L'approccio da noi scelto, l'analisi fattoriale confermativa multilivello, rappresenta sicuramente un approccio relativamente nuovo e molto più adatto a valutare l'invarianza di misurazione nelle indagini caratterizzate da un ampio numero di unità di secondo livello (Rabe-Hesketh et al. 2004; Muthén 1990).

Innanzitutto, è stato stimato un modello di analisi fattoriale confermativa ad un livello. Tale modello ci ha consentito di mostrare come la struttura fattoriale a sette dimensioni ipotizzata teoricamente fosse sostenibile anche a livello empirico. In tal modo innanzitutto è stato mostrato come la coesione sociale sia un concetto multidimensionale: infatti, un modello con un sola dimensione ha mostrato un pessimo adattamento ai dati (RMSEA = .099, CFI = .437)³⁶. Stabilito che il modello a sette dimensioni ha mostrato un buon adattamento ai dati, abbiamo proseguito la nostra analisi specificando un modello di analisi fattoriale confermativa multilivello. L'aspetto più importante riferito a tale modello è il fatto che abbia mostrato isomorfismo (per approfondimenti si veda Adamopoulos 2008; Fontaine 2008) nella struttura fattoriale di primo e secondo livello. Ciò sottintende che la struttura a sette dimensioni della coesione sociale sia sostenibile non solo a livello individuale ma anche a livello aggregato. In altre parole, il modello multilivello della coesione sociale è equivalente fra i livelli (*cross-level equivalent*) dal momento che presenta la stessa struttura fattoriale passando da un livello ad un altro. In tal modo, abbiamo illustrato come, non solo il costrutto di coesione sociale sia un costrutto multidimensionale, ma anche un costrutto multilivello. Attraverso il Modello-2 è stato ulteriormente mostrato come oltre ad avere la stessa struttura fattoriale a livello-1 e a livello-2, i sette fattori latenti di primo livello presentino relazioni con i primi indicatori uguali alle relazioni che i fattori di secondo livello hanno con gli indicatori di secondo livello. In altre parole, possiamo dire che i fattori di primo e di secondo livello hanno fra di loro la stessa interpretazione (i coefficienti lambda sono uguali fra i livelli). In aggiunta, al fine di testare l'invarianza di misurazione abbiamo specificato il Modello-3 che ci ha consentito di concludere come tale modello di coesione sociale sia, inoltre, anche invariante

³⁵ Per diversi non si intende alternativi, abbiamo visto come il Modello-1 2 e 3 siano uguali e si differenzino solo per alcuni parametri che venivano liberati o fissati al fine di testare diverse ipotesi.

³⁶ Tale modello non è stato presentato data la sua irrilevanza.

all'interno di ogni Paese considerato, ovvero il modello di coesione sociale è esattamente lo stesso in ognuno dei 29 Stati.

Infine, attraverso la stima del modello di analisi fattoriale confermativa multilivello di secondo ordine siamo stati in grado di mostrare la presenza di un fattore generale isomorfo di secondo ordine a livello aggregato o di società (Paese) che abbiamo denominato *coesione sociale* e definito come *inclusione sociale* a livello individuale.

Ciò ci permette di affermare che la coesione sociale è un concetto fondato certamente sugli atteggiamenti e i comportamenti degli individui ma allo stesso tempo esercita un'influenza da un punto di vista aggregato. I Paesi con maggior coesione sono il frutto di condizioni aggregate che producono negli individui atteggiamenti e comportamenti favorevoli alla generazione di inclusione e nei confronti del senso di appartenenza facilitando ed incrementando interazioni positive e, a loro volta, tali interazioni interpersonali operano al fine di mantenere e riprodurre tali condizioni positive (si veda anche Friedkin 2004). La coesione sociale deve essere considerata un concetto ascrivibile ad un livello aggregato in quanto è inopportuno affermare il grado di coesione di un singolo individuo. A livello individuale, infatti, riteniamo che vi sia un altro meccanismo in gioco, ossia l'*inclusione sociale*. Tale concetto, nella nostra teorizzazione, è il corrispettivo concetto isomorfo che opera a livello individuale della coesione sociale che dà conto degli atteggiamenti e dei comportamenti individuali (quelli che operano per riprodurre le condizioni positive) come i livelli di fiducia interpersonale, la partecipazione, la densità delle relazioni sociali. Tali atteggiamenti e comportamenti sono, come detto, influenzati anche da meccanismi macro, nel nostro caso ricondotti alla *coesione sociale*.

Dal momento che i pesi fattoriali sono equiparati fra i due livelli, la rilevanza degli effetti macro e di quelli micro possono essere "misurati" e comparati. Abbiamo visto che il livello macro nella spiegazione degli atteggiamenti e comportamenti dà conto del 24% della variabilità nel costrutto latente *fiducia interpersonale*, del 14,6% nel fattore *densità delle relazioni sociali*, del 5% nella dimensione *supporto sociale*, del 13,5% nel fattore *partecipazione*, del 13% in *apertura*, quasi il 29% nel costrutto *fiducia nelle istituzioni* e, infine, il 31% nella *legittimità delle istituzioni*. Inoltre, in riferimento al costrutto generale *coesione sociale*, abbiamo visto come quasi il 35% della variabilità del fenomeno sia attribuibile al contesto macro (Paese) in cui il soggetto risiede. Al di là dei semplici numeri, capiamo come l'effetto individuale – o effetto micro – sia certamente più ampio di quello macro, tuttavia quest'ultimo rimane comunque rilevante e non può essere in nessun modo trascurato. Attraverso il modello di analisi fattoriale multilivello di secondo ordine abbiamo mostrato come la coesione sociale sia un concetto multidimensionale di secondo ordine. In termini sostantivi ciò si traduce nel fatto che la coesione sociale non esercita in maniera diretta i suoi effetti sugli atteggiamenti e i comportamenti individuali, piuttosto la coesione sociale è un concetto più generale che produce effetti sulle dimensioni costituenti individuate che a loro volta impattano sulla sfera individuale.

Lo stesso modello multilivello della coesione sociale ha mostrato la sua validità all'interno di tutti e 29 gli Stati considerati rivelando come i meccanismi della coesione sociale operino alla stessa maniera in tutti i Paesi considerati.

L'invarianza di misurazione riscontrata nel modello ci permette anche di confrontare le medie riferite ai costrutti latenti fra i diversi Paesi oggetto di indagine che altrimenti sarebbero state non confrontabili in caso di non invarianza del modello. Per calcolare i punteggi fattoriali riferiti sia ai sette costrutti della coesione sociale presi singolarmente che al costrutto generale *coesione sociale* di secondo ordine abbiamo specificato un ulteriore modello di analisi fattoriale confermativa – esattamente identico al modello di secondo ordine prima presentato – ricorrendo però ad uno stimatore di massima verosimiglianza (incrementato in Mplus con l'acronimo MLR, si veda il capitolo

2). La motivazione che ci ha condotto alla sostituzione del metodo di stima è riferita esclusivamente all'impossibilità di calcolare i punteggi fattoriali ricorrendo ad uno stimatore WLSMV (Muthén e Muthén 1998-2012). In tab.3.10 riportiamo tutti i punteggi fattoriali relativi sia alle sette sotto-dimensioni della coesione sociale sia al costrutto generale per i 29 Paesi oggetto della nostra indagine (i Paesi sono ordinati in ordine decrescente in base al livello di coesione sociale). Ricordiamo che i punteggi fattoriali sono valori standardizzati con media zero e deviazione standard uguale ad 1.

Tab.3.10 – Punteggi fattoriali riferiti alla coesione sociale e alle sue sette sotto-dimensioni

Paese	Coesione	Fiducia interpersonale	Intensità relazioni sociali	Supporto sociale	Apertura	Partecipazione	Fiducia nelle istituzioni	Legittimità delle istituzioni
Denmark	1.876	2.092	1.452	1.786	1.036	0.807	1.889	1.594
Norway	1.798	1.791	1.448	1.679	1.039	1.650	1.797	1.624
Switzerland	1.758	0.973	1.031	1.300	0.949	1.692	1.717	2.015
Finland	1.587	1.602	0.499	0.090	1.340	0.227	1.557	1.729
Sweden	1.456	1.267	1.794	1.232	1.712	0.570	1.458	1.271
Netherlands	1.430	1.056	1.344	0.410	0.744	1.460	1.503	0.974
Belgium	0.891	0.124	0.532	-0.336	-0.070	0.154	0.884	1.135
Germany	0.544	0.259	0.640	1.368	0.837	1.633	0.527	0.518
Iceland	0.542	1.460	1.202	1.132	1.761	1.869	0.499	0.399
United Kingdom	0.367	0.639	0.283	-0.114	-0.538	0.831	0.365	0.374
Israel	0.114	0.177	0.408	0.540	-0.408	0.116	0.043	0.515
France	0.013	-0.111	1.100	0.265	-0.412	0.435	0.008	0.028
Estonia	0.008	0.400	-1.387	-0.866	-0.063	-0.911	0.033	-0.075
Hungary	-0.112	-0.268	-2.015	-0.049	-0.570	-1.247	0.002	-0.647
Ireland	-0.132	0.451	-0.280	-0.238	0.207	0.590	-0.145	-0.156
Lithuania	-0.340	-0.004	-1.762	-0.363	0.051	-1.207	-0.376	-0.087
Russian Federation	-0.446	-0.493	-0.870	-2.076	-2.087	-0.851	-0.327	-0.857
Cyprus	-0.501	-1.532	-0.972	1.223	-2.543	-0.488	-0.498	-0.240
Czech Republic	-0.521	-0.574	-0.528	-1.825	-1.231	-0.780	-0.561	-0.051
Slovakia	-0.563	-1.110	-0.919	-1.082	-0.970	-0.324	-0.550	-0.392
Poland	-0.832	-0.914	-0.697	0.170	1.319	-0.968	-0.889	-0.571
Slovenia	-0.840	-0.282	-0.395	0.427	-0.225	-0.355	-0.864	-0.823
Italy	-0.943	-0.578	-0.160	-0.943	-0.117	0.397	-0.973	-0.789
Albania	-0.973	-1.665	-0.795	-0.644	1.040	-0.849	-0.971	-0.894
Spain	-0.980	0.005	0.532	0.214	0.671	0.975	-1.044	-0.907
Kosovo	-1.203	-1.482	-0.353	-0.173	-1.110	-0.780	-1.186	-1.211
Portugal	-1.240	-1.227	0.933	-1.129	-0.932	-0.849	-1.232	-1.241
Ukraine	-1.455	-0.699	-0.617	-1.055	-0.666	-1.183	-1.444	-1.570
Bulgaria	-1.466	-1.833	-0.562	0.726	0.133	-1.486	-1.439	-1.663

Notiamo che i Paesi con una maggiore dotazione in termini di coesione sociali siano quelli scandinavi insieme alla Svizzera e ai Paesi Bassi. L'Italia si colloca in una posizione molto bassa, in linea con Paesi come la Spagna, l'Albania, la Slovenia e la Polonia. Se teniamo conto non solo dell'indice generale di coesione sociale ma analizziamo singolarmente le sotto-dimensioni che lo compongono, osserviamo

come la coesione sociale si configuri in maniera non omogenea all'interno dei singoli Paesi con le singole sotto-dimensioni che assumono peso variabile a seconda del Paese di riferimento.

4. Le dimensioni della Coesione Sociale, la frammentazione etnica e la disuguaglianza: un modello ad equazioni strutturali multilivello

La specificazione di un modello di analisi fattoriale confermativa multilivello (MLCFA) non costituisce necessariamente l'ultimo *step* di un'indagine empirica. Attraverso l'inserimento all'interno del modello di MLCFA si può specificare un Modello ad equazioni strutturali multilivello (Asparouhov e Muthén 2007; Rabe-Hesketh, Skrondal e Pickles 2004; Rabe-Hesketh, Skrondal e Pickles 2007; Rabe-Hesketh, Skrondal e Zheng 2012; Mehta e Neale 2005; Depaoli e Clifton 2015; Preacher, Zyphur e Zhang 2010, Hox 2013). Le covariate possono essere aggiunte al modello generale soltanto dopo che è stato raggiunto un modello di misurazione che mostra un buon adattamento ai dati. Viceversa, risulterebbe senza nessun senso logico specificare delle relazioni fra costrutti che non rilevano quello che dovrebbero rilevare. All'interno di un modello ad equazioni strutturali multilivello (MSEM), le covariate possono essere introdotte sia a livello-1 della gerarchia sia a livello-2 che ad entrambi i livelli. Il modello, inoltre può contenere termini di interazione fra i livelli, o includere ipotesi di mediazione o moderazione. Rispetto ad un modello di regressione multilivello classico che prevede che la variabile dipendente debba essere necessariamente di primo livello, uno dei principali vantaggi consiste nella possibilità di inserire nel modello disegni di ricerca che includano variabili dipendenti di secondo livello (per approfondimenti sui MSEM si veda il capitolo 2).

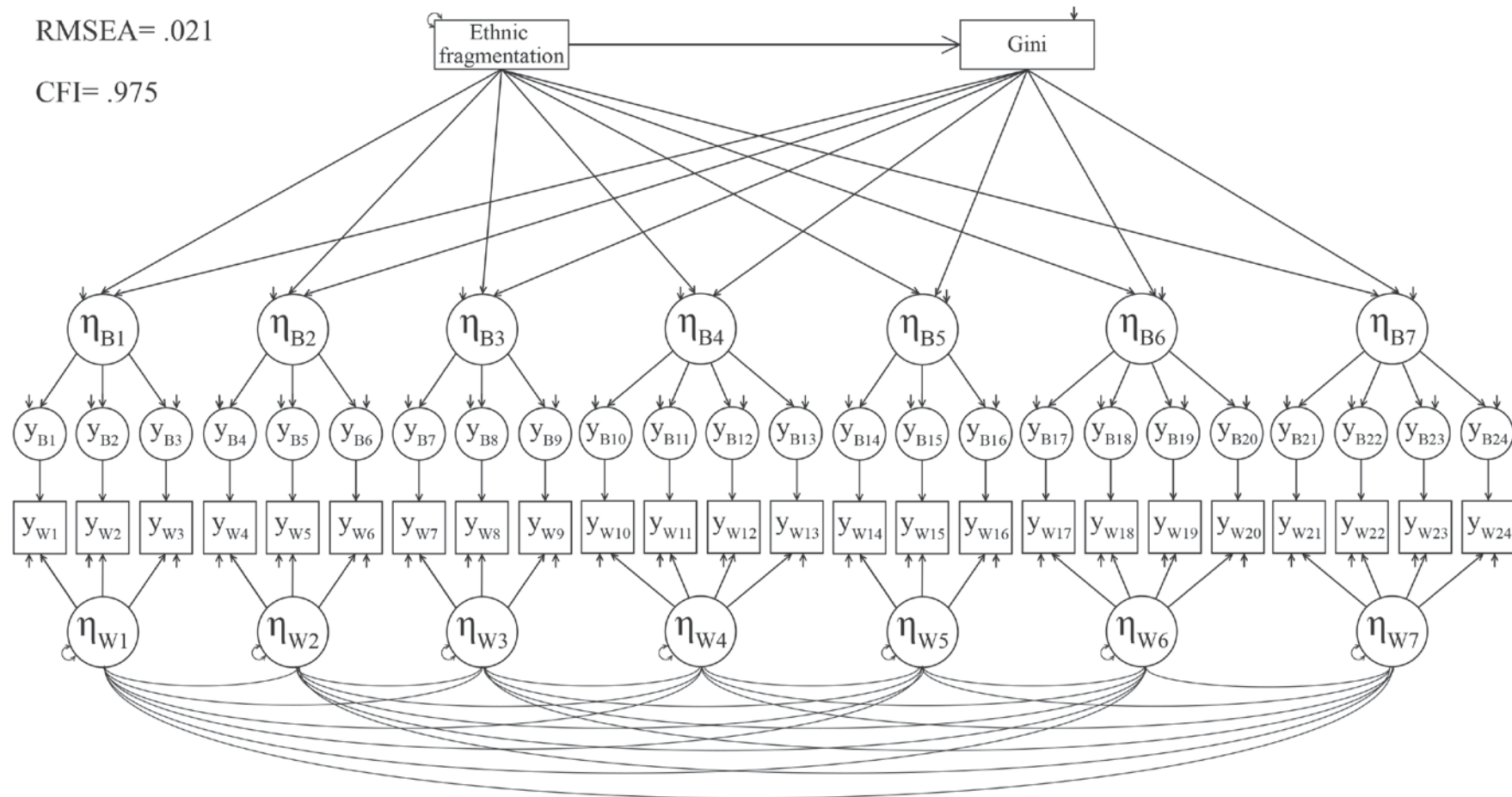
Nel lavoro qui presentato, è stato stimato un modello ad equazioni strutturali multilivello con una struttura che include un predittore di secondo livello, un mediatore di secondo livello e una variabile dipendente di secondo livello. In modo più specifico, l'obiettivo consiste nel valutare l'effetto della *frammentazione etnica*³⁷ sui sette costrutti che compongono la coesione sociale valutandone quanto tale effetto sia un effetto diretto o mediato dalla dimensione della disuguaglianza rilevata attraverso il coefficiente di Gini (fig.4.1)

Come hanno mostrato numerose indagini, la *frammentazione etnica* di un Paese può esercitare un effetto negativo sulla coesione sociale. Secondo Putnam (2007), la diversità è strettamente connessa ad un più basso livello di capitale sociale registrato – al crescere della diversità il capitale sociale diminuisce. Hawes and Rocha (2011), sostengono che la diversità esercita un effetto negativo sulla sfera afferente alla fiducia sociale, così come Hero (2007) afferma che la diversità etnica negli Stati Uniti è negativamente connessa al capitale sociale e alla cultura civica.

Tuttavia, non vi è accordo unanime su tali aspetti e sulle relazioni individuate in tali studi. Infatti, una ricerca condotta all'interno dei Paesi Bassi ha mostrato come l'effetto della diversità etnica sulla coesione sociale sia un effetto che può variare in intensità e direzione in base alle dimensioni della coesione sociale scelte per rilevare tale concetto (Tolsma, van der Meer and Gesthuizen 2009). In Canada, all'inverso, è stato mostrato che quanto più le città si caratterizzavano per diversità tanto più si registravano nella popolazione livelli alti di fiducia (Kazemipur 2006).

³⁷ La frammentazione etnica rileva il grado di eterogeneità da un punto di vista etnico, religioso e linguistico. La variabile si riferisce alla probabilità che due soggetti estratti a caso in un dato Paese non appartengano allo stesso gruppo etnico, religioso o linguistico (Alesina, Devleeschauwer, Easterly, Kurlat e Wacziarg 2003).

Fig.4.1 – Modello multilivello ad equazioni strutturali



Tenendo in debito conto i risultati appena illustrati, la mia ipotesi generale si riferisce alla possibilità che la *frammentazione etnica* non eserciti un effetto diretto sui costrutti della coesione sociale, piuttosto è la dimensione della disuguaglianza (rilevata tramite *coefficiente di Gini*) che, interagendo con la *frammentazione etnica*, esercita un effetto negativo sulle diverse dimensioni della coesione sociale. La disuguaglianza, all'interno del nostro impianto teorico, riveste un ruolo fondamentale. Infatti, la deprivazione è direttamente collegata a meccanismi che conducono a forme, più o meno severe, di esclusione, intaccando la coesione sociale. Più precisamente, la carenza di risorse economiche conduce ad una mancanza nella sfera delle capacità (*capabilities*) del soggetto (Sen 1992). A tal proposito, Sen distingue fra *functioning* e *capabilities*. Il *functioning* è "an achievement of a person: what she or he manages to do or be. It reflects, as it were, a part of the 'state' of that person" (Sen, 1985, p.10). Una *capability* è, invece, la capacità che il soggetto ha di raggiungere un dato *functioning* in termini di "essere e fare". Sen concepisce l'esistenza di una persona come una combinazione di "*doings e beings*" (*functionings*) e valuta il suo benessere nei termini della libertà di scegliere tra varie combinazioni (*capabilities*) di questi *functionings*. Quindi, una scarsità di risorse materiali comporta una mancanza nelle capacità di un soggetto di raggiungere una serie di "funzionamenti" che vanno ad intaccare aspetti – come la fiducia interpersonale, la partecipazione sociale e politica, la fiducia nelle istituzioni – che costituiscono elementi fondamentali nella partecipazione attiva alla vita sociale contribuendo a creare coesione.

Pertanto, cercheremo di stabilire se e quanto le relazioni fra *frammentazione etnica* e i sette costrutti della coesione sociale siano mediate dalla dimensione della disuguaglianza.

Per testare tale ipotesi, è stato stimato un modello ad equazioni strutturali multilivello presentato in fig.4.1. La variabile indipendente è la *frammentazione etnica*, il mediatore è il *coefficiente di Gini*, infine, le variabili dipendenti sono i sette fattori latenti della coesione sociale.

Il modello presenta un ottimo adattamento ai dati. L'indice RMSEA pari a .021 e l'indice CFI (.975) sono, rispettivamente, al di sotto e al di sopra dei valori soglia consigliati in letteratura indicanti un eccellente adattamento del modello. Alcuni coefficienti strutturali, con l'eccezione di *legittimità delle istituzioni* regredito su Gini ($p < .001$), *fiducia istituzionale* regredito su Gini ($p = .003$), *supporto sociale* regredito su Gini ($p = .014$) e *fiducia interpersonale* regredito su Gini ($p = .005$), non sono significativi. Tale condizione potrebbe essere dovuta all'ampiezza campionaria di secondo livello abbastanza esigua (i coefficienti che stiamo analizzando sono coefficienti di secondo livello pertanto il campione è quello riferito ai 29 Paesi e non ai 54672 intervistati) e dalla mancanza di potenza statistica che ne consegue. È plausibile che in un campione di secondo livello più ampio, tenendo conto anche della grandezza delle stime ottenute, i coefficienti risulterebbero statisticamente significativi. Si porrà pertanto maggiore attenzione alla grandezza delle stime dei coefficienti strutturali piuttosto che agli aspetti legati alla significatività statistica.

Per rendere più chiara l'esposizione dei risultati, precisiamo che in un modello di mediazione classico abbiamo generalmente tre coefficienti strutturali o detti *path* che coinvolgono la variabile dipendente, indipendente e quella mediatrice (Baron e Kenny 1986). Il coefficiente strutturale denominato "path a" si riferisce al coefficiente di regressione relativo all'effetto esercitato dalla variabile indipendente X sul mediatore M. Nel nostro modello il path a è riferito al coefficiente relativo all'effetto della *frammentazione etnica* esercitato sul mediatore *coefficiente di Gini*. Il path b, invece, si riferisce all'effetto condizionato esercitato dalla variabile mediatrice M sulla variabile indipendente Y quando X è incluso nel modello. Nel nostro caso specifico, il path b rappresenta l'effetto del *coefficiente di Gini* sui sette fattori latenti della coesione sociale al netto dell'effetto della *frammentazione etnica*. Infine, il path c1 si riferisce all'effetto condizionato esercitato dalla variabile indipendente X sulla variabile

dipendente Y quando il mediatore M è incluso nel modello. Nel caso specifico, il path c1 è l'effetto diretto della *frammentazione etnica* sui sette costrutti latenti della coesione sociale valutato al netto dell'influenza esercitata dal *coefficiente di Gini*. L'effetto indiretto, che è l'effetto esercitato da X su Y attraverso il mediatore M è uguale ad $a*b$ o $c-c1$, dove il path c è l'effetto totale di X su Y in un semplice modello bivariato.

Facendo riferimento alla tab.4.1, possiamo notare come in un semplice modello ad equazioni strutturali multilivello bivariato la *frammentazione etnica* eserciti un effetto negativo sulle dimensioni della coesione sociale con i coefficienti di regressione che variano da $b = -.115$ a $b = -.391$ (si veda tab.4.1: Effetto totale: frammentazione etnica path c).

Aggiungendo al modello la dimensione della disuguaglianza, ossia il *coefficiente di Gini*, i coefficienti riferiti a tali relazioni si modificano. L'effetto negativo esercitato dalla *frammentazione etnica* sulla *legittimità delle istituzioni* risulta essere pari quasi a zero ($b = .07$). L'effetto esercitato della *frammentazione etnica* è pertanto un effetto principalmente indiretto che produce la sua influenza attraverso il *coefficiente di Gini* ($b = -.187$). Al netto dell'influenza esercitata dalla *frammentazione etnica*, il *coefficiente di Gini* mostra un ampio effetto negativo sulla dimensione *legittimità delle istituzioni* ($b = -.662$).

Il costrutto *fiducia nelle istituzioni* è influenzato negativamente dalla *frammentazione etnica* in un semplice modello bivariato. Ma nel modello di mediazione, con l'aggiunta del aspetto relativo alle disuguaglianze, tale relazione scompare quasi del tutto ($b = -.08$). Nuovamente, la relazione è mediata dal *coefficiente di Gini* e la *frammentazione etnica* esercita la sua influenza sulla *fiducia nelle istituzioni* in un maniera solo indiretta influenzando positivamente il *coefficiente di Gini* che impatta sul costrutto in esame. Il *coefficiente di Gini*, inoltre, si configura come un forte predittore negativo della *fiducia nelle istituzioni* (il coefficiente condizionato è pari a $b = -.587$).

Tab.4.1 – Coefficienti strutturali standardizzati nel modello di mediazione ad equazioni strutturali multilivello

	Effetto totale: Frammentazione etnica (path c)	Effetto diretto: Frammentazione etnica (path c1)	Effetto indiretto: Frammentazione etnica (path a*b)	Gini (path b)	Frammentazione etnica (path a)
Gini					0.282
Legittimità delle istituzioni	-0.115	0.071	-0.187	-0.662	
Fiducia nelle istituzioni	-0.254	-0.089	-0.166	-0.587	
Apertura	-0.140	-0.079	-0.061	-0.216	
Partecipazione	-0.285	-0.193	-0.092	-0.327	
Supporto sociale	-0.391	-0.269	-0.122	-0.432	
Densità delle relazioni sociali	-0.349	-0.218	-0.130	-0.461	
Fiducia interpersonale	-0.221	-0.073	-0.148	-0.526	

In una società in cui sussiste maggiore diversità di tipo etnico ci si potrebbe semplicisticamente aspettare una maggiore apertura in riferimento alla sfera che ingloba i fenomeni legati all'immigrazione, ma nel modello da noi specificato avviene invece il contrario. Se pur piccolo, la

frammentazione etnica esercita un effetto negativo sulla dimensione latente *apertura* ($b = -.140$). Tale effetto, con l'introduzione della variabile mediatrice, si riduce approssimandosi quasi a zero ($b = -.07$), tuttavia non possiamo affermare nulla in riferimento a tale relazione in quanto i coefficienti chiamati in causa risultano essere esigui (si veda tab.4.1). Possiamo solo notare come il *coefficiente di Gini* eserciti un'influenza negativa di medio-bassa intensità sul costrutto *apertura*. Notiamo, inoltre, come tale effetto sia quello più piccolo comparato agli altri effetti che *Gini* esercita sulle altre dimensioni. Ciò potrebbe indicare che sul costrutto relativo alla disponibilità nei confronti dell'immigrazione potrebbero esercitare la loro influenza altre dimensioni afferenti maggiormente ad aspetti legati alla sfera culturale piuttosto che meramente economica.

Gli effetti di mediazione esercitati dalla variabile *coefficiente di Gini* sulle relazioni fra la *frammentazione etnica* e i costrutti della coesione sociale visti fino ad ora sembrano non interessare nella stessa misura le dimensioni relative alla *partecipazione*, al *supporto sociale* e alla *densità delle relazioni sociali*. Infatti, notiamo come l'effetto diretto relativo alle succitate relazioni rimanga ancora ampio (tab.4.1). Gli effetti di mediazione relativi al *supporto sociale* e alla *densità delle relazioni sociali* ancora sussistono ($b = -.122$ per *supporto sociale* e $-.13$ per *densità delle relazioni sociali*, mentre l'effetto indiretto per la *partecipazione* è prossimo allo zero) ma sono effetti parziali in quanto si registrano stime dei coefficienti strutturali relativi all'effetto diretto abbastanza ampi ($b = -.269$ per *supporto sociale* e $b = -.218$ per *densità delle relazioni sociali*). Possiamo concludere che la *frammentazione etnica*, pur avendo una porzione che risulta mediata dal *coefficiente di Gini*, esercita comunque un'influenza negativa diretta sui costrutti relativi al *supporto sociale* e alla *densità delle relazioni sociali*.

L'effetto prodotto dalla *frammentazione etnica* torna ad essere un effetto principalmente mediato dal *coefficiente di Gini* relativamente all'ultimo costrutto preso in considerazione, ossia la *fiducia interpersonale*. L'effetto totale negativo esercitato dalla *frammentazione etnica* sulla *fiducia interpersonale* è pari ad un coefficiente di $-.221$, ma tale effetto è principalmente mediato dal *coefficiente di Gini* in quanto l'effetto diretto della *frammentazione etnica* sulla *fiducia interpersonale*, dopo l'introduzione della variabile mediatrice, si riduce quasi a zero ($b = -.07$) mentre l'effetto indiretto è stimabile in $b = -.148$.

La dimensione della disuguaglianza rilevata attraverso il *coefficiente di Gini* sembra essere un buon mediatore delle relazioni fra la *frammentazione etnica* e i costrutti relativi alla coesione sociale anche tenendo presente i coefficienti relativi al R^2 (tab.4.2). La porzione di varianza spiegata (o riprodotta) relativa alle sette dimensioni della coesione sociale cresce in modo considerevole passando da un modello bivariato ad un modello di mediazione con l'introduzione della variabile *coefficiente di Gini*. R^2 aumenta da un 5% di varianza spiegata fino al 30% relativamente al costrutto *legittimità delle istituzioni*, dal 12% al 32% per il fattore latente *fiducia nelle istituzioni*, dal 15% fino al 32,5% per la dimensione *apertura*, dal 8% al 18% per *partecipazione*, dal 2% al 6% per il fattore *supporto sociale*, dal 6,5% fino al 38% per la dimensione *densità delle relazioni sociali*, e infine, passa dal 13% fino al 42% per il costrutto *fiducia interpersonale*.

Tab.4.2 – Stime dell' R^2

	R^2 modello bivariato	R^2 modello con GINI
<i>Legittimità delle istituzioni</i>	0.049	0.303
<i>Fiducia nelle istituzioni</i>	0.122	0.317

<i>Apertura</i>	0.153	0.324
<i>Partecipazione</i>	0.082	0.180
<i>Supporto sociale</i>	0.020	0.063
<i>Densità delle relazioni sociali</i>	0.065	0.381
<i>Fiducia interpersonale</i>	0.013	0.417

Sintetizzando, possiamo affermare che la *frammentazione etnica* non sempre esercita un'influenza negativa sulle dimensioni che costituiscono la coesione sociale. Abbiamo visto come tale relazione è spesso mediata da elementi afferenti ad un ambito di tipo economico declinato nel nostro caso in termini di disuguaglianze di reddito fra i cittadini di un Paese. Si può concludere sostenendo che, all'interno di una società, ampi livelli di diversità non necessariamente impattano negativamente sulla coesione sociale mettendone a rischio l'ordine sociale, piuttosto è il grado di disuguaglianza che sussiste fra i soggetti che può produrre fenomeni di erosione della coesione sociale.

Conclusioni

L'obiettivo principale di questo lavoro di tesi sono stati quelli di fornire una definizione operativa multilivello della coesione sociale e valutare successivamente gli effetti a livello aggregato che i fattori riferibili alla diversità e alla disuguaglianza esercitano sulla coesione di un Paese. Per tale motivo, innanzitutto si è proceduto a realizzare un modello di analisi fattoriale confermativa multilivello e successivamente abbiamo specificato un modello completo ad equazioni strutturali multilivello.

Prima dell'analisi empirica, si è provveduto ad effettuare un'attenta ed esaustiva ricognizione della letteratura scientifica presente in materia. In seguito a tale fase, abbiamo presentato la nostra prospettiva teorica relativa alla coesione sociale fornendo uno schema concettuale con l'obiettivo di costituire una base teorica di partenza volta a coadiuvare future indagini sulla coesione sociale nella fase di operativizzazione e rilevazione.

Sulla base dello schema teorico si è proceduto a sviluppare la nostra proposta di definizione operativa del concetto di coesione. Sono state individuate sette sotto-dimensioni che si pongono come fattori costituenti del costrutto in esame.

L'analisi fattoriale multilivello ha confermato empiricamente il nostro schema teorico a sette sotto-dimensioni e il modello statistico ha mostrando un ottimo adattamento. In particolare, attraverso un modello di analisi fattoriale multilivello di secondo ordine abbiamo messo in evidenza l'esistenza di due fattori generali, rispettivamente di primo e di secondo livello, responsabili delle covariazioni fra i sette fattori di primo ordine. Il fattore macro è stato denominato *coesione sociale*, il fattore micro (individuale) *inclusione sociale*. Questo elemento costituisce sicuramente uno dei principali punti di forza del lavoro in quanto si è riusciti ad individuare un singolo fattore più generale – di ordine maggiore – capace di sintetizzare le sette sotto-dimensioni. Infatti, fra tutti gli studi presenti in materia, nessuno è stato in grado di conseguire un tale risultato, fallendo nell'individuazione di un unico indice di coesione. Un ulteriore punto di notevole importanza teorica, si riferisce al fatto che il modello multilivello di coesione sociale così come specificato risulta essere invariante fra i 29 Paesi analizzati sia in riferimento all'aspetto relativo alla misurazione che a quello strutturale (all'interno dei 29 Paesi, la struttura e le relazioni fra indicatori e costrutti latenti risultano essere le medesime, così come l'intensità degli effetti esercitati dai costrutti sulle variabili osservate). Ciò ci permette di affermare che i meccanismi responsabili della coesione sociale risultano essere gli stessi all'interno dei Paesi analizzati. Pertanto il modello di coesione sociale presentato ha la caratteristica fondamentale di prescindere dalle specificità di ogni singola cultura e acquistare valenza generale all'interno dell'Europa. Nonostante le differenze culturali, il modello di coesione risulta valido all'interno di ogni contesto analizzato.

La coesione sociale è un concetto ascrivibile ad un livello esclusivamente aggregato in quanto è illogico valutare il grado di coesione di un singolo individuo. A livello individuale, il meccanismo in gioco si riferisce all'*inclusione sociale*. Tale concetto, nella nostra teorizzazione, è il corrispettivo concetto isomorfo della coesione sociale che opera a livello individuale.

Il costrutto macro e il costrutto micro (coesione ed inclusione sociale) danno conto pertanto dei fenomeni di superficie che osserviamo e allo stesso tempo risultano responsabili di alcune delle dimensioni fondamentali dell'agire umano come la fiducia negli altri individui, la partecipazione, la disponibilità ad accettare soggetti immigrati e la densità delle relazioni sociali.

La coesione sociale è un concetto che emerge dagli atteggiamenti e i comportamenti individuali ma contemporaneamente esercita un effetto di tipo aggregato. Pertanto, in riferimento al meccanismo di generazione della coesione, possiamo affermare che una maggiore dotazione di coesione all'interno

di un Paese è il frutto di condizioni aggregate le quali producono negli individui atteggiamenti e comportamenti favorevoli alla generazione di inclusione e senso di appartenenza facilitando ed incrementando le interazioni positive. A loro volta, tali interazioni interpersonali operano al fine di mantenere e riprodurre tali condizioni positive.

Successivamente, è stata indagata l'influenza in termini di effetti macro che la diversità e la disuguaglianza esercitano sulla coesione sociale. Il modello di mediazione ad equazioni strutturali multilivello ha evidenziato come la diversità etnica all'interno di un Paese non necessariamente influisce negativamente sulla coesione sociale. Tra la diversità e la coesione si frappongono altre dimensioni che possono mediare tale relazione. Nel caso da noi presentato, la relazione è mediata dalle disuguaglianze di reddito sussistenti fra individui.

Concludendo, in una società, la diversità non necessariamente riduce i livelli di coesione sociale. Tale conclusione è in linea con le considerazioni teoriche di Durkheim (1983) secondo il quale, in una società moderna, la diversità costituirebbe il meccanismo di produzione di coesione laddove, invece, la coesione generata da condivisione dello stesso sistema valoriale, degli stessi orizzonti interpretativi e dalle stesse rappresentazioni sarebbe tipico di una società arcaica.

Appendice statistica

ESS Countries and sample size

Country	Sample Size	%	Country	Sample Size	%
Albania	1201	2.2	Israel	2508	4.6
Belgium	1869	3.4	Iceland	752	1.4
Bulgaria	2259	4.1	Italy	960	1.8
Switzerland	1493	2.7	Lithuania	2109	3.9
Cyprus	1116	2.0	Netherlands	1845	3.4
Czech Republic	2009	3.7	Norway	1624	3.0
Germany	2958	5.4	Poland	1898	3.5
Denmark	1650	3.0	Portugal	2151	3.9
Estonia	2380	4.4	Russian Federation	2484	4.5
Spain	1889	3.5	Sweden	1847	3.4
Finland	2197	4.0	Slovenia	1257	2.3
France	1968	3.6	Slovakia	1847	3.4
United Kingdom	2286	4.2	Ukraine	2178	4.0
Hungary	2014	3.7	Kosovo	1295	2.4
Ireland	2628	4.8	Total	54672	100

1 – Interpersonal trust

A3 – Most people can be trusted or you can't be too careful

	Frequency	Valid Percent	Cumulative Percent
You can't be too careful	3609	6.6	6.6
1	2494	4.6	11.2
2	3876	7.1	18.3
3	5750	10.6	28.9
4	5294	9.7	38.6
5	10474	19.2	57.8
6	6089	11.2	69.0
7	8096	14.9	83.9
8	6167	11.3	95.2
9	1621	3.0	98.2
Most people can be trusted	983	1.8	100.0
Total	54453	100.0	
Refusal	10		
Don't know	180		
No answer	30		
Total	220		
Total	54673		

A4 – Most people try to take advantage of you, or try to be fair

	Frequency	Valid Percent	Cumulative Percent
Most people try to take advantage of me	1931	3.6	3.6
1	1557	2.9	6.4
2	2833	5.2	11.7
3	4515	8.3	20.0

4	4812	8.9	28.9
5	11108	20.5	49.5
6	6373	11.8	61.3
7	9312	17.2	78.5
8	7834	14.5	93.0
9	2312	4.3	97.2
Most people try to be fair	1500	2.8	100.0
Total	54087	100.0	
Refusal	8		
Don't know	545		
No answer	33		
Total	586		
Total	54673		

A5 – Most of the time people helpful or mostly looking out for themselves

	Frequency	Valid Percent	Cumulative Percent
People mostly look out for themselves	2884	5.3	5.3
1	2388	4.4	9.7
2	4285	7.9	17.6
3	6108	11.2	28.8
4	6023	11.1	39.9
5	11071	20.4	60.3
6	6651	12.2	72.5
7	7337	13.5	86.0
8	5142	9.5	95.4
9	1500	2.8	98.2
People mostly try to be helpful	982	1.8	100.0
Total	54371	100.0	
Refusal	9		
Don't know	241		
No answer	52		
Total	302		
Total	54673		

2 – Density of social relations

C2 – How often socially meet with friends, relatives or colleagues

	Frequency	Percent	Valid Percent	Cumulative Percent
Never	1236	2,3	2,3	2,3
Less than once a month	5196	9,5	9,6	11,8
Once a month	5601	10,2	10,3	22,1
Several times a month	10629	19,4	19,6	41,7
Once a week	9162	16,8	16,9	58,6
Several times a week	14039	25,7	25,8	84,4
Every day	8488	15,5	15,6	100,0
Total	54351	99,4	100,0	
Refusal	18	,0		
Don't know	276	,5		
No answer	28	,1		
Total	322	,6		

Total	54673	100,0
-------	-------	-------

C3 – How many people with whom you can discuss intimate and personal matters

	Frequency	Percent	Valid Percent	Cumulative Percent
None	2977	5,4	5,5	5,5
1	9709	17,8	18,1	23,6
2	12280	22,5	22,8	46,4
3	11900	21,8	22,1	68,6
4-6	12633	23,1	23,5	92,1
7-9	2316	4,2	4,3	96,4
10 or more	1938	3,5	3,6	100,0
Total	53753	98,3	100,0	
Refusal	82	,1		
Don't know	696	1,3		
No answer	142	,3		
Total	920	1,7		
Total	54673	100,0		

C4 – Take part in social activities compared to others of same age

	Frequency	Percent	Valid Percent	Cumulative Percent
Much less than most	7167	13,1	13,5	13,5
Less than most	13675	25,0	25,7	39,2
About the same	23657	43,3	44,5	83,8
More than most	6956	12,7	13,1	96,9
Much more than most	1661	3,0	3,1	100,0
Total	53116	97,2	100,0	
Refusal	21	,0		
Don't know	1512	2,8		
No answer	24	,0		
Total	1557	2,8		
Total	54673	100,0		

3 – Social support

D21 – Feel people in local area help one another

	Frequency	Percent	Valid Percent	Cumulative Percent
Not at all	2533	4,6	4,7	4,7
1	2996	5,5	5,6	10,3
2	6099	11,2	11,4	21,7
3	11485	21,0	21,5	43,2
4	12544	22,9	23,4	66,6
5	11439	20,9	21,4	88,0
A great deal	6418	11,7	12,0	100,0
Total	53514	97,9	100,0	

Refusal	27	,0
Don't know	1111	2,0
No answer	21	,0
Total	1159	2,1
Total	54673	100,0

D29 – Feel appreciated by people you are close to

	Frequency	Percent	Valid Percent	Cumulative Percent
Not at all	148	,3	,3	,3
1	144	,3	,3	,5
2	338	,6	,6	1,2
3	764	1,4	1,4	2,6
4	1168	2,1	2,2	4,7
5	3669	6,7	6,8	11,5
6	4357	8,0	8,1	19,6
7	8465	15,5	15,7	35,2
8	14387	26,3	26,6	61,9
9	10189	18,6	18,8	80,7
Completely	10433	19,1	19,3	100,0
Total	54062	98,9	100,0	
Refusal	24	,0		
Don't know	558	1,0		
No answer	29	,1		
Total	611	1,1		
Total	54673	100,0		

D36 – Receive help and support from people you are close to

	Frequency	Percent	Valid Percent	Cumulative Percent
Not at all	598	1,1	1,1	1,1
1	642	1,2	1,2	2,3
2	1485	2,7	2,7	5,0
3	3945	7,2	7,3	12,3
4	7600	13,9	14,0	26,3
5	16679	30,5	30,8	57,1
Completely	23209	42,5	42,9	100,0
Total	54158	99,1	100,0	
Refusal	136	,2		
Don't know	348	,6		
No answer	31	,1		
Total	515	,9		
Total	54673	100,0		

4 – Participation

	Yes	No	Refusal	Don't know	No answer
B12 Worked in political party or action group last 12 months	3.7	96.0	.1	.2	.1
B13 Worked in another organisation or association last 12 months	13.1	86.6	.0	.2	.1
B15 Signed petition last 12 months	19.1	80.3	.0	.4	.1

D1 –Involved in work for voluntary or charitable organisations, how often past 12 months

	Frequency	Valid Percent	Cumulative Percent
At least once a week	3431	6.3	6.3
At least once a month	3603	6.7	13.0
At least once every three months	2565	4.7	17.7
At least once every six months	3184	5.9	23.6
Less often	6715	12.4	36.0
Never	34669	64.0	100.0
Total	54167	100.0	
Refusal	29		
Don't know	400		
No answer	77		
Total	506		
Total	54673		

5 – Openness

B32 – Immigration bad or good for country's economy

	Frequency	Percent	Valid Percent	Cumulative Percent
Bad for the economy	3715	6,8	7,2	7,2
1	2342	4,3	4,5	11,7
2	3828	7,0	7,4	19,1
3	5210	9,5	10,1	29,2
4	4980	9,1	9,6	38,8
5	11843	21,7	22,9	61,7
6	5459	10,0	10,5	72,2
7	6074	11,1	11,7	83,9
8	4767	8,7	9,2	93,1
9	1465	2,7	2,8	96,0
Good for the economy	2086	3,8	4,0	100,0
Total	51769	94,7	100,0	
Refusal	60	,1		
Don't know	2803	5,1		
No answer	41	,1		
Total	2904	5,3		
Total	54673	100,0		

B33 – Country's cultural life undermined or enriched by immigrants

	Frequency	Percent	Valid Percent	Cumulative Percent
Cultural life undermined	2754	5,0	5,3	5,3

1	1824	3,3	3,5	8,8
2	3031	5,5	5,8	14,7
3	4059	7,4	7,8	22,5
4	3993	7,3	7,7	30,2
5	10403	19,0	20,1	50,2
6	5452	10,0	10,5	60,7
7	7310	13,4	14,1	74,8
8	6943	12,7	13,4	88,2
9	2669	4,9	5,1	93,4
Cultural life enriched	3446	6,3	6,6	100,0
Total	51884	94,9	100,0	
Refusal	42	,1		
Don't know	2714	5,0		
No answer	33	,1		
Total	2789	5,1		
Total	54673	100,0		

B34 – Immigrants make country worse or better place to live

	Frequency	Percent	Valid Percent	Cumulative Percent
Worse place to live	2995	5,5	5,8	5,8
1	1972	3,6	3,8	9,6
2	3580	6,5	6,9	16,6
3	4747	8,7	9,2	25,8
4	5138	9,4	10,0	35,7
5	14526	26,6	28,2	63,9
6	5298	9,7	10,3	74,2
7	5456	10,0	10,6	84,8
8	4401	8,0	8,5	93,3
9	1516	2,8	2,9	96,2
Better place to live	1938	3,5	3,8	100,0
Total	51567	94,3	100,0	
Refusal	69	,1		
Don't know	2990	5,5		
No answer	47	,1		
Total	3106	5,7		
Total	54673	100,0		

6 – Institutional trust

	B2 Trust in country's parliament	B3 Trust in the legal system	B5 Trust in politicians	B6 Trust in political parties
No trust at all	15.3	10.9	19.8	19.0
1	6.8	5.5	9.4	9.6
2	9.5	7.6	12.2	12.1

3	11.6	9.5	12.5	12.8
4	9.9	9.2	10.7	10.8
5	15.8	14.4	14.6	15.1
6	9.1	9.4	8.4	8.1
7	9.4	11.3	6.3	5.9
8	6.4	11.5	2.8	2.7
9	1.9	5.2	.7	.6
Complete trust	1.5	2.8	.5	.6
Refusal	.1	.1	.1	.1
Don't know	2.5	2.6	1.9	2.4
No answer	.1	.1	.1	.1
Total	100.0	100.0	100.0	100.0

7 – Legitimacy of institutions

	B22 How satisfied with the national government	B23 How satisfied with the way democracy works in country
Extremely dissatisfied	13.7	6.7
1	7.3	3.7
2	10	6.6
3	12.3	9.3
4	11.2	10.0
5	15.4	16.3
6	10.8	11.9
7	10.4	14.8
8	6	13.1
9	1.7	5.0
Extremely satisfied	1.1	2.5
Total	100	100.0

	B24 State of education in country nowadays	B25 State of health services in country nowadays
Extremely bad	3.9	6.1
1	2.8	4.8
2	5.5	7.7
3	8.7	10.0
4	9.8	9.9
5	15.7	13.5
6	13.3	11.5
7	16.5	14.0
8	14.8	14.2
9	6.0	5.4
Extremely good	3.1	3.0
Total	100.0	100.0

DEMOGRAPHICS VARIABLES

1 – Gender

	Frequency	Percent	Valid Percent	Cumulative Percent
Male	24929	45,6	45,6	45,6
Female	29727	54,4	54,4	100,0
Total	54656	100,0	100,0	
No answer	17	,0		
Total	54673	100,0		

2 – Age

N valid	54540
N missing	133
Mean	48,31
Std. Deviation	18,592
Min.	15
Max.	103
Qaurtile 25	33
Qaurtile 50	48
Qaurtile 75	63

3 – Education

	Freq.	%	Valid %	Cumulative %
ES-ISCED I , less than lower secondary	5897	10,8	10,8	10,8
ES-ISCED II, lower secondary	9819	18,0	18,0	28,8
ES-ISCED IIIb, lower tier upper secondary	7404	13,5	13,6	42,4
ES-ISCED IIIa, upper tier upper secondary	12123	22,2	22,2	64,7
ES-ISCED IV, advanced vocational, sub-degree	7336	13,4	13,5	78,1
ES-ISCED V1, lower tertiary education, BA level	5019	9,2	9,2	87,3
ES-ISCED V2, higher tertiary education, >= MA level	6711	12,3	12,3	99,7
Other	186	,3	,3	100,0
Total	54495	99,7	100,0	
Refusal	30	,1		
Don't know	75	,1		
No answer	73	,1		
Total	178	,3		
Total	54673	100,0		

4 – Years of full-time education completed

N valid	54233
N missing	440
Mean	12,54
Std. Deviation	4,030

Min.	0
Max.	51
Qaurtile 25	10,00
Qaurtile 50	12,00
Qaurtile 75	15,00

5 – Partner's highest level of education, ES - ISCED

	Frequency	Percent	Valid Percent	Cumulative Percent
ES-ISCED I , less than lower secondary	3024	5.5	9.6	9.6
ES-ISCED II, lower secondary	4882	8.9	15.5	25.2
ES-ISCED IIIb, lower tier upper secondary	5026	9.2	16.0	41.2
ES-ISCED IIIa, upper tier upper secondary	6575	12.0	20.9	62.1
ES-ISCED IV, advanced vocational, sub-degree	4393	8.0	14.0	76.1
ES-ISCED V1, lower tertiary education, BA level	3242	5.9	10.3	86.4
ES-ISCED V2, higher tertiary education, >= MA level	4089	7.5	13.0	99.5
Other	168	.3	.5	100.0
Total	31399	57.4	100.0	
Not applicable	22603	41.3		
Refusal	144	.3		
Don't know	288	.5		
No answer	239	.4		
Total	23274	42.6		
Total	54673	100.0		

6 – Father's highest level of education, ES - ISCED

	Frequency	Percent	Valid Percent	Cumulative Percent
ES-ISCED I , less than lower secondary	16072	29.4	32.8	32.8
ES-ISCED II, lower secondary	9067	16.6	18.5	51.2
ES-ISCED IIIb, lower tier upper secondary	7295	13.3	14.9	66.1
ES-ISCED IIIa, upper tier upper secondary	6092	11.1	12.4	78.5
ES-ISCED IV, advanced vocational, sub-degree	4190	7.7	8.5	87.1
ES-ISCED V1, lower tertiary education, BA level	2172	4.0	4.4	91.5
ES-ISCED V2, higher tertiary education, >= MA level	3818	7.0	7.8	99.3
Other	357	.7	.7	100.0
Total	49063	89.7	100.0	
Refusal	287	.5		
Don't know	5214	9.5		
No answer	109	.2		
Total	5610	10.3		
Total	54673	100.0		

7 – Mother's highest level of education, ES - ISCED

	Frequency	Percent	Valid Percent	Cumulative Percent
ES-ISCED I , less than lower secondary	18325	33.5	35.9	35.9
ES-ISCED II, lower secondary	11188	20.5	21.9	57.8
ES-ISCED IIIb, lower tier upper secondary	5819	10.6	11.4	69.2
ES-ISCED IIIa, upper tier upper secondary	6898	12.6	13.5	82.7
ES-ISCED IV, advanced vocational, sub-degree	3966	7.3	7.8	90.4
ES-ISCED V1, lower tertiary education, BA level	2162	4.0	4.2	94.7
ES-ISCED V2, higher tertiary education, >= MA level	2544	4.7	5.0	99.6
Other	182	.3	.4	100.0
Total	51084	93.4	100.0	
Refusal	227	.4		
Don't know	3264	6.0		
No answer	98	.2		
Total	3589	6.6		
Total	54673	100.0		

8 – Employment Status

	Frequency	Percent	Valid Percent	Cumulative Percent
Paid work	25760	47,1	47,6	47,6
Education	4704	8,6	8,7	56,3
Unemployed, looking for job	2978	5,4	5,5	61,8
Unemployed, not looking for job	1155	2,1	2,1	63,9
Permanently sick or disabled	1291	2,4	2,4	66,3
Retired	12819	23,4	23,7	90,0
Community or military service	90	,2	,2	90,2
Housework, looking after children, others	4796	8,8	8,9	99,0
Other	520	1,0	1,0	100,0
Total	54113	99,0	100,0	
Refusal	39	,1		
Don't know	105	,2		
No answer	416	,8		
Total	560	1,0		
Total	54673	100,0		

9 – Ever unemployed and seeking work for a period more than three months

	Frequency	Percent	Valid Percent	Cumulative Percent
Yes	16153	29,5	29,7	29,7
No	38147	69,8	70,3	100,0
Total	54300	99,3	100,0	
Refusal	57	,1		
Don't know	225	,4		

No answer	91	,2
Total	373	,7
Total	54673	100,0

10 – Any period of unemployment and work seeking lasted 12 months or more

	Frequency	Percent	Valid Percent	Cumulative Percent
Yes	8168	14,9	50,8	50,8
No	7914	14,5	49,2	100,0
Total	16082	29,4	100,0	
Not applicable	38429	70,3		
Refusal	5	,0		
Don't know	87	,2		
No answer	70	,1		
Total	38591	70,6		
Total	54673	100,0		

11 – Partner's main activity

	Frequency	N %
Partner doing last 7 days: paid work	18449	31.1
Partner doing last 7 days: education	662	1.1
Partner doing last 7 days: unemployed, actively looking for job	1518	2.6
Partner doing last 7 days: unemployed, not actively looking for job	637	1.1
Partner doing last 7 days: permanently sick or disabled	743	1.3
Partner doing last 7 days: retired	7779	13.1
Partner doing last 7 days: community or military service	36	0.1
Partner doing last 7 days: housework, looking after children, others	5875	9.9
Partner doing last 7 days: other	552	0.9
Partner doing last 7 days: don't know	107	0.2
Partner doing last 7 days: not applicable	22603	38.1
Partner doing last 7 days: refusal	55	0.1
Partner doing last 7 days: no answer	286	0.5
Total	59302	100

11 – Household's total net income, all sources

	Frequency	Percent	Valid Percent	Cumulative Percent
1st decile	4979	9,1	11,3	11,3
2nd decile	5302	9,7	12,1	23,4
3rd decile	5014	9,2	11,4	34,8
4th decile	4793	8,8	10,9	45,7
5th decile	4619	8,4	10,5	56,2
6th decile	4277	7,8	9,7	65,9

7th decile	4175	7,6	9,5	75,4
8th decile	3829	7,0	8,7	84,1
9th decile	3452	6,3	7,8	91,9
10th decile	3541	6,5	8,1	100,0
Total	43981	80,4	100,0	
Refusal	6211	11,4		
Don't know	4342	7,9		
No answer	139	,3		
Total	10692	19,6		
Total	54673	100,0		

12 – Feeling about household's income nowadays

	Frequency	Percent	Valid Percent	Cumulative Percent
Living comfortably on present income	12785	23,4	23,7	23,7
Coping on present income	23120	42,3	42,8	66,5
Difficult on present income	12157	22,2	22,5	89,0
Very difficult on present income	5915	10,8	11,0	100,0
Total	53977	98,7	100,0	
Refusal	126	,2		
Don't know	490	,9		
No answer	80	,1		
Total	696	1,3		
Total	54673	100,0		

Bibliografia

- Adamopoulos, J. (2008). On the entanglement of culture and individual behavior. *Multilevel analysis of individuals and cultures*. In F. J. Van de Vijver, D. A. Van Hemert and Y. H. Poortinga (2008). *Multilevel analysis of individuals and cultures* (pp. 27-62). Lawrence Erlbaum Associates, Taylor & Francis Group, LLC.
- Alesina, A., Devleeschauwer, A., Easterly, W., Kurlat, S., & Wacziarg, R. (2003). Fractionalization. *Journal of Economic growth*, 8(2), 155-194.
- Asparouhov, T., & Muthén, B. (2003). Full-information maximum-likelihood estimation of general two-level latent variable models with missing data: A technical report. *Los Angeles: Muthén & Muthén*.
- Asparouhov, T., and Muthén, B. (2007). Computationally efficient estimation of multilevel high-dimensional latent variable models. In *proceedings of the 2007 JSM meeting in Salt Lake City, Utah, Section on Statistics in Epidemiology* (pp. 2531-2535).
- Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6), 1173.
- Barr, R., & Dreeben, R. (1983) *How Schools Work*. Chicago: University of Chicago Press.
- Bauer, D. J. (2003). Estimating multilevel linear models as structural equation models. *Journal of Educational and Behavioral Statistics*, 28(2), 135-167.
- Beauvais, C., e Jenson, J. (2002). Social cohesion: Updating the state of the research. CPRN Discussion Paper No. F | 22 (Canadian Policy Research Networks, Ottawa).
- Bentler, P. M., & Chou, C. P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16(1), 78-117.
- Bentler, P. M. (1992). On the fit of models to covariances and methodology to the Bulletin. *Psychological bulletin*, 112(3), 400.
- Berger-Schmitt, R. (2000). Social cohesion as an aspect of the quality of societies: concept and measurement. *Eureporting Working Paper n°14 (Centre for Survey Research and Methodology, Mannheim)*.
- Berger-Schmitt, (2002). Considering social cohesion in quality of life assessments: concept and measurement, *Social Indicators Research* 58: 403–428, 2002.
- Berger-Schmitt, R. e H.-H. Noll (2000), 'Conceptual framework and structure of a European system of social indicators', *EuReporting Working Paper No. 9 (Centre for Survey Research and Methodology (ZUMA), Social Indicators Department, Mannheim)*.
- Bernard, P. (1999). Social cohesion: A dialectical critique of a quasi-concept. Paper SRA-491, Ottawa: Strategic Research and Analysis Directorate. Department of Canadian Heritage.
- Blalock, H. M. (1967). Causal inferences in nonexperimental research. University of North Carolina Press

- Bollen, K. A., e Hoyle, R. H. (1990). Perceived cohesion: A conceptual and empirical examination. *Social forces*, 479-504.
- Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons, Inc.
- Bollen, K. A., & Long, J. S. (1993). *Testing structural equation models*. Beverly Hills, Sage.
- Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective* (Vol. 467). John Wiley & Sons.
- Boudon, R. (1977). *Effets pervers et ordre social* (Vol. 1). Presses universitaires de France.
- Bovaird J.A., (2007) Multilevel Structural Equation Models for Contextual Factors, in Little, T. D., Bovaird, J. A., & Card, N. A. (Eds.). (2007). *Modeling contextual effects in longitudinal studies*. Routledge.
- Byrne, B. M., & van De Vijver, F. J. (2010). Testing for measurement and structural equivalence in large-scale cross-cultural studies: Addressing the issue of nonequivalence. *International Journal of Testing*, 10(2), 107-132.
- Chan, D. (1998). Functional relations among constructs in the same content domain at different levels of analysis: A typology of composition models. *Journal of applied psychology*, 83(2), 234.
- Chan, J., To, H. P. e Chan, E. (2006). Reconsidering social cohesion: Developing a definition and analytical framework for empirical research. *Social indicators research*, 75(2), 273-302.
- Cheung, M. W. L., and Au, K. (2005). Applications of multilevel structural equation modeling to cross-cultural research. *Structural Equation Modeling*, 12(4), 598-619.
- Cieciuch, J., Davidov, E., Schmidt, P., Algesheimer, R., and Schwartz, S. H. (2014). Comparing results of an exact vs. an approximate (Bayesian) measurement invariance test: a cross-country illustration with a scale to measure 19 human values. *Frontiers in psychology*, 5.
- Collins, R. (1988). *Theoretical sociology*. Harcourt College Pub.
- Council of Europe, Objective 1997 - The intergovernmental programme of activities of the Council of Europe. Strasbourg: Council of Europe, 1997.
- Curran, P. J. (2003). Have multilevel models been structural equation models all along?. *Multivariate Behavioral Research*, 38(4), 529-569.
- Davidov, E., Meuleman, B., Cieciuch, J., Schmidt, P., and Billiet, J. (2014). Measurement equivalence in cross-national research. *Annual Review of Sociology*, 40: 55–75.
- Depaoli, S., & Clifton, J. P. (2015). A Bayesian approach to multilevel structural equation modeling with continuous and dichotomous outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, 22(3), 327-351.
- Dickes, P., Valentova, M., and Borsenberger, M. (2010). Construct validation and application of a common measure of social cohesion in 33 European countries. *Social Indicators Research*, 98(3), 451-473.

- Dickes, P., and Valentova, M. (2013). Construction, validation and application of the measurement of social cohesion in 47 European countries and regions. *Social indicators research*, 113(3), 827-846.
- du Toit, H.C. & du Toit, M. (2008). Multilevel structural equation modeling. In J. de Leeuw & E. Meijer (Ed.), *Handbook of Multilevel Analysis* (pp. 435-479). New York, NY: Springer.
- Duhaime, G., Searles, E., Usher, P. J., Myers, H., and Frechette, P. (2004). Social cohesion and living conditions in the Canadian Arctic: From theory to measurement. *Social Indicators Research*, 66(3), 295-318.
- Duncan, O. D. (1966). Path analysis: Sociological examples. *American journal of Sociology*, 1-16.
- Durkheim, E. (1893). De la division du travail social: étude sur l'organisation des sociétés supérieures. Presses Universitaires de France.
- Durkheim, E. (1897). *Le suicide: étude de sociologie*. F. Alcan.
- Dyer, N. G., Hanges, P. J., & Hall, R. J. (2005). Applying multilevel confirmatory factor analysis techniques to the study of leadership. *The Leadership Quarterly*, 16(1), 149-167.
- Ekström, J. (2011). A generalized definition of the polychoric correlation coefficient. *Department of Statistics, UCLA*.
- European Committee for Social Cohesion (2004). A new Strategy for Social Cohesion, Strasbourg, Council of Europe.
- European Commission (2014). Sixth Report on economic, social and territorial cohesion, Bruxelles.
- European Union, First Report on Economic and Social Cohesion 1996. Luxembourg: Office for Official Publications of the European Communities, 1996.
- Fischer, R., (2008). Multilevel Approaches in Organizational Settings: Opportunities, Challenges and Implications for Cross-Cultural Research. In F. J. Van de Vijver, D. A Van Hemert and Y. H Poortinga (2008). *Multilevel analysis of individuals and cultures* (pp. 65-92). Lawrence Erlbaum Associates, Taylor & Francis Group, LLC.
- Fontaine, J. R. (2008). Traditional and multilevel approaches in cross-cultural research: An integration of methodological frameworks. In F. J. Van de Vijver, D. A Van Hemert and Y. H Poortinga (2008). *Multilevel analysis of individuals and cultures* (pp. 65-92). Lawrence Erlbaum Associates, Taylor & Francis Group, LLC.
- Friedkin, N. E. (2004). Social cohesion. *Annual Review of Sociology*, 30: 409-425
- Geddes, M. (1995). The Role of Partnerships in Promoting Social Cohesion, European Foundation for the Improvement of Living and Working Conditions Working Paper No. WP/95/38/EN, Dublin.
- Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. Univ of California Press.

- Goldberger, A. S. (1972). Structural equation methods in the social sciences. *Econometrica: Journal of the Econometric Society*, 979-1001.
- Goldstein, H., & McDonald, R. P. (1988). A general model for the analysis of multilevel data. *Psychometrika*, 53(4), 455-467.
- Granovetter, M. S. (1973). The strength of weak ties. *American journal of sociology*, 1360-1380.
- Hawes, D. P., & Rocha, R. R. (2011). Social Capital, Racial Diversity, and Equity Evaluating the Determinants of Equity in the United States. *Political Research Quarterly*, 64(4), 924-937.
- Hero, Rodney E. (2007). *Racial Diversity and Social Capital*. New York: Cambridge University Press.
- Hofstede, G. (1980). *Culture's consequences: International differences in work-related values*. Beverly Hills, CA: Sage.
- House of Commons ODPM: housing, planning, local government and the Regions committee (2004). *Social Cohesion, Sixth Report of Session 2003_04*, vol. 1 Report, London, Office of the Deputy Prime Minister, The Stationery Office.
- Hox, J. J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum.
- Hox, J. (2010). *Multilevel analysis: Techniques and applications*. Routledge.
- Hox, J. (2013). Multilevel Regression and Multilevel Structural Equation Modeling. In T. D. Little, (2013). *The Oxford handbook of quantitative methods in psychology* (Vol. 2) (pp. 281-294). Oxford University Press.
- Hox, J. J., & Maas, C. J. (2001). The accuracy of multilevel structural equation modeling with pseudobalanced groups and small samples. *Structural equation modeling*, 8(2), 157-174.
- Hox, J. J., Maas, C. J., & Brinkhuis, M. J. (2010). The effect of estimation method and sample size in multilevel structural equation modeling. *Statistica Neerlandica*, 64(2), 157-170.
- Hu, L. T., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1-55.
- Jak, S., Oort, F. J., and Dolan, C. V. (2013). A test for cluster bias: Detecting violations of measurement invariance across clusters in multilevel data. *Structural Equation Modeling: A Multidisciplinary Journal*, 20(2), 265-282.
- Jeannotte, M. S. (2003). *Social cohesion: Insights from Canadian research*. Gatineau: Department of Canadian Heritage.
- Jeannotte, M. S. (2000). *Social cohesion around the world: An international comparison of definitions and issues*, Strategic Research and Analysis, Ottawa: Canadian Heritage.
- Jenson, J., (1998). *Mapping social cohesion: The state of Canadian research*. Paper SRA-321, Family Network, Canadian Policy Research Networks Inc.
- Jöreskog, K. G. (1999). How large can a standardized coefficient be?. *The help-file of the LISREL program*.

- Jöreskog, K. G., e Sörbom, D. (1979). *Advances in factor analysis and structural equation models*, Cambridge, Mass., Abt Books.
- Jöreskog, K. G., e Sörbom, D. (1988). *LISREL 7: A guide to the program and applications*, Chicago Ill., Spss Inc.
- Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. *Structural Equation Modeling*, 8(3), 325-352.
- Kaplan, D., & Kreisman, M. B. (2000). On the validation of indicators of mathematics education using TIMSS: An application of multilevel covariance structure modeling. *International Journal of Educational Policy, Research, and Practice*, 1(2), 217-242.
- Kazemipur, A. (2006). A Canadian exceptionalism? Trust and diversity in Canadian cities. *Journal of International Migration and Integration/Revue de l'integration et de la migration internationale*, 7(2), 219-240.
- Kline, R. B. (2011). *Principles and practice of Structural Equation Modelling, Third Edition*, Londra e New York, The Guilford Press.
- Krull, J. L., & MacKinnon, D. P. (1999). Multilevel mediation modeling in group-based intervention studies. *Evaluation Review*, 23(4), 418-444.
- Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate behavioral research*, 36(2), 249-277.
- Lazarsfeld, P. F. (1954). A conceptual introduction to latent structure analysis. *Mathematical thinking in the social sciences*, 1, 349-387.
- Lazarsfeld, P. F., e Menzel, H. (1961). On the relation between individual and collective properties. *Complex organizations: A sociological reader*, 422-440.
- Le Bon, G., (1895). *Psychologie des foules*, Alcan.
- Li, F., Duncan, T. E., Harmer, P., Acock, A., & Stoolmiller, M. (1998). Analyzing measurement models of latent variables through multilevel confirmatory factor analysis and hierarchical linear modeling approaches. *Structural Equation Modeling: A Multidisciplinary Journal*, 5(3), 294-306.
- Liang, J., & Bentler, P. M. (2004). An EM algorithm for fitting two-level structural equation models. *Psychometrika*, 69(1), 101-122.
- Lockwood, D. (1964). Social Integration and System integration, in *Explorations in Social Change*, G. K. Zollschan and W. Hirsch (eds). London: Routledge.
- Lockwood, D. (1992). *Solidarity and Schism: The Problem of Disorder in Durkheimian and Marxist Sociology*, Oxford, Clarendon Press.
- Lockwood, D. (1999). Civic integration and social cohesion. In Gough, I. (1999.) *Capitalism and Social Cohesion* (pp. 63-84). Palgrave Macmillan UK.
- MacCallum, R. C., Browne, M. W., and Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, 1(2), 130.

- Marradi, A. (1981). Misurazione e scale: qualche riflessione e una proposta. *Quaderni di sociologia*, 29, 595-639.
- McDonald, R. P. (1994). The bilevel reticular action model for path analysis with latent variables. *Sociological Methods and Research*, 22, 399-413.
- Mehta, P. D., & Neale, M. C. (2005). People are variables too: multilevel structural equations modeling. *Psychological methods*, 10(3), 259.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58(4), 525-543.
- Merton, R. K. (1949). *Social theory and social structure*, New York, Free Press.
- Michalski, W., Miller, R., and Stevens, B. (1997). Economic Flexibility and Societal Cohesion in the Twenty-First Century: An Overview of the Issues and Key Points of the Discussion, in *Societal Cohesion and Globalising Economy – What Does the Future Hold?* Paris: OECD, pp. 7-25.
- Moerbeek, M. (2004). The consequence of ignoring a level of nesting in multilevel analysis. *Multivariate Behavioral Research*, 39(1), 129-149.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered, categorical and continuous latent variable indicators. *Psychometrika*, 49(1), 115-132.
- Muthén, B. O. (1989). Latent variable modeling in heterogeneous populations. *Psychometrika*, 54(4), 557-585.
- Muthén, B. (1990). Mean and covariance structure analysis of hierarchical data. Los Angeles, CA: UCLA.
- Muthén, B. O. (1991). Multilevel factor analysis of class and student achievement components. *Journal of Educational Measurement*, 28(4), 338-354.
- Muthén, B. O. (1994). Multilevel covariance structure analysis. *Sociological methods & research*, 22(3), 376-398.
- Muthén, B., and A. Satorra. 1989. "Multilevel Aspects of Varying Parameters for the Factor Analysis of Nonnormal Likert Variables." In *Multilevel Analysis of Educational Data*, edited by R. D. Bock, 87-99. San Diego: Academic Press.
- Muthén, B. O., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological methodology*, 267-316.
- Muthén, B. O., e Muthén, L. K. (1998-2012). *Mplus User's guide*. Los Angeles, CA: Muthén & Muthén.
- Noll, H., (2002), Toward a European System of Social Indicators: Theoretical framework and System Architecture, in Michael R. Hagerty, Joachim Vogel, Valerie Møller, *Assessing Quality of Life and Living Conditions to Guide* (Social Indicators Research Series), 2002, Springer.
- Parsons, T. (1949). *The structure of social action* (Vol. 491). New York: Free Press.

- Pearson, K. (1900). Mathematical contributions to the theory of evolution. VII. On the correlation of characters not quantitatively measurable. *Philos. Trans. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci.*, 195, 1–47.
- Pituch, K. A., Stapleton, L. M., & Kang, J. Y. (2006). A comparison of single sample and bootstrap methods to assess mediation in cluster randomized trials. *Multivariate Behavioral Research*, 41(3), 367-400.
- Pornprasertmanit, S., Lee, J., & Preacher, K. J. (2014). Ignoring clustering in confirmatory factor analysis: some consequences for model fit and standardized parameter estimates. *Multivariate behavioral research*, 49(6), 518-543.
- Preacher, K. J., Zyphur, M. J., & Zhang, Z. (2010). A general multilevel SEM framework for assessing multilevel mediation. *Psychological methods*, 15(3), 209.
- Preacher, K. J., Zhang, Z., & Zyphur, M. J. (2011). Alternative methods for assessing mediation in multilevel data: The advantages of multilevel SEM. *Structural Equation Modeling*, 18(2), 161-182.
- Putnam, R. D. (2001). *Bowling alone: The collapse and revival of American community*. Simon and Schuster.
- Putnam, R. D. (2007). E pluribus unum: Diversity and community in the twenty-first century the 2006 Johan Skytte Prize Lecture. *Scandinavian political studies*, 30(2), 137-174.
- Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). Generalized multilevel structural equation modeling. *Psychometrika*, 69(2), 167-190.
- Rabe-Hesketh, S., Skrondal, A., & Zeng, X. (2007). *Multilevel structural equation modeling* (pp. 209-227), in S. Y. Lee (Ed.). *Handbook on Structural Structural Equation Models*. Amsterdam: Elsevier.
- Rabe-Hesketh, S., Skrondal, A., & Zeng, X. (2012). *Multilevel Structural Equation Modeling*, in *Handbook of structural equation modeling*, Hoyle, R. H. (Ed.). (2012).. Guilford Press.
- Rajulton, F., Ravanera, Z. R., and Beaujot, R. (2007). Measuring social cohesion: An experiment using the Canadian national survey of giving, volunteering, and participating. *Social Indicators Research*, 80(3), 461-492.
- Raudenbush, S. W., Rowan, B., & Kang, S. J. (1991). A multilevel, multivariate model for studying school climate with estimation via the EM algorithm and application to US high-school data. *Journal of Educational and Behavioral Statistics*, 16(4), 295-330.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (Vol. 1). Sage.
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15, 351-357.
- Rovine, M. J., & Molenaar, P. C. M. (1998). A nonstandard method for estimating a linear growth model in LISREL. *International Journal of Behavioral Development*, 22, 453–473.
- Rovine, M. J., & Molenaar, P. C. (2000). A structural modeling approach to a multilevel random coefficients model. *Multivariate Behavioral Research*, 35(1), 51-88.

- Rovine, M. J., & Molenaar, P. C. M. (2001). A structural equations modeling approach to the general linear mixed model. In L. M. Collins and A. G. Sayer (Eds.), *New methods for the analysis of change* (pp. 65–96). Washington, DC: American Psychological Association.
- Ryu, E., & West, S. G. (2009). Level-specific evaluation of model fit in multilevel structural equation modeling. *Structural Equation Modeling*, 16(4), 583-601.
- Sen, A., (1985). *Commodities and Capabilities*, Oxford: Elsevier Science Publishers.
- Sen, Amartya K. (1992), *Inequality Reexamined*, Oxford: Clarendon Press.
- Satorra, A., & Bentler, P.M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C.C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.
- Selig, J. P., Card, N. A., and Little, T. D. (2008). Latent variable structural equation modeling in cross-cultural research: Multigroup and multilevel approaches. In F. J. Van de Vijver, D. A. Van Hemert and Y. H. Poortinga (2008). *Multilevel analysis of individuals and cultures* (pp. 93-119). Lawrence Erlbaum Associates, Taylor & Francis Group, LLC.
- Snijder, T., and Bosker, R. (1999). *Multilevel analysis. An introduction to basic and advanced modelling*. London, UK: SAGE Publications.
- Social, Health and Family Affairs Committee (1998). Fighting social exclusion and strengthening social cohesion in Europe. Strasbourg: Parliamentary Assembly of the Council of Europe.
- Spearman, Charles (1904) "'General Intelligence', Objectively Determined and Measured", *American Journal of Psychology*, 15: 201-293.
- Stanat, P. e Ludtke, O. (2008). Multilevel issues in international large-scale assessment studies on student performance. In F. J. Van de Vijver, D. A. Van Hemert and Y. H. Poortinga (2008). *Multilevel analysis of individuals and cultures* (pp. 93-119). Lawrence Erlbaum Associates, Taylor & Francis Group, LLC.
- Steiger, J. H., and Lind, J. C. (1980). Statistically based tests for the number of common factors. In *annual meeting of the Psychometric Society, Iowa City, IA* (Vol. 758).
- Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. Routledge
- The European Task Force on Cultural and Development (1997). In from the margins – A contribution to the debate on Culture and Development in Europe. Strasbourg: Council of Europe.
- Tolsma, J., van der Meer, T., & Gesthuizen, M. (2009). The impact of neighbourhood and municipality characteristics on social cohesion in the Netherlands. *Acta Politica*, 44(3), 286–313.
- Triandis, H. C. (2001). Individualism-collectivism and personality. *Journal of Personality*, 69, 907–924.
- Van de Vijve, F. J., Van Hemert, D. A., Poortinga, Y. H., (2008). Conceptual Issues in Multilevel Models. In F. J. Van de Vijver, D. A. Van Hemert and Y. H. Poortinga (2008). *Multilevel analysis of individuals and cultures* (pp. 93-119). Lawrence Erlbaum Associates, Taylor & Francis Group, LLC.

- Van de Vijver, F. J. R., & Poortinga, Y. H. (2002). Structural equivalence in multilevel research. *Journal of Cross-Cultural Psychology*, 33, 141–156.
- Weber, M., (1922) 1978. *Economy and Society: An Outline of Interpretive Sociology*, Berkley, CA: U. California Press.
- Whelan, C. T., & Maître, B. (2005). Economic vulnerability, multidimensional deprivation and social cohesion in an enlarged European community. *International Journal of Comparative Sociology*, 46(3), 215-239.
- Yuan, K.H. & Bentler, P.M. (2000). Three likelihood-based methods for mean and covariance structure analysis with nonnormal missing data. In M.E. Sobel & M.P. Becker (eds.), *Sociological Methodology 2000* (pp. 165-200). Washington, D.C.: ASA.
- Yuan, K. H., & Hayashi, K. (2005). On Muthén's maximum likelihood for two-level covariance structure models. *Psychometrika*, 70(1), 147-167.
- Zhang, Z., Zyphur, M. J., & Preacher, K. J. (2009). Testing multilevel mediation using hierarchical linear models problems and solutions. *Organizational Research Methods*, 12(4), 695-719.