

Università degli Studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Magistrale in
Scienze Statistiche



**VALIDAZIONE DEL QUESTIONARIO DI VALUTAZIONE
DELLA DIDATTICA DA PARTE DEGLI STUDENTI
DELL'UNIVERSITA' DEGLI STUDI DI PADOVA:
APPROCCIO TRADIZIONALE E A CLASSI LATENTI**

Relatore Prof.ssa Francesca Bassi
Dipartimento di Scienze Statistiche

Laureanda: Cipriana Baldinazzo
Matricola N. 1061026

Anno Accademico 2015/2016

INDICE

INTRODUZIONE	1
CAPITOLO 1	3
SVILUPPO E VALIDAZIONE DI UNA SCALA DI MISURA: APPROCCIO TRADIZIONALE	3
1.1 Processo di sviluppo di una scala di misura	3
1.2 Affidabilità	8
1.2.1 Indici di correlazione.....	8
1.2.2 Analisi fattoriale.....	9
1.2.3 Alpha di Cronbach	10
1.2.4 Indici di coerenza interna basati sul metodo <i>split-half</i>	12
1.2.5 Procedura split-half sample	12
1.3 Validità.....	13
1.3.1 Validità di contenuto	13
1.3.2 Validità di criterio	14
1.3.3 Validità di costrutto.....	14
CAPITOLO 2	15
PROGETTO PRODID E QUESTIONARIO RIVOLTO AGLI STUDENTI DELL'UNIVERSITA' DEGLI STUDI DI PADOVA.....	15
2.1 Il progetto PRODID	15
2.2 Il questionario.....	16
2.3 I dati	18
2.3.1 Statistiche descrittive.....	18
CAPITOLO 3	21
VALIDAZIONE DEL QUESTIONARIO	21
3.1 Verifica dell'affidabilità interna.....	22
3.1.1 Analisi di correlazione	22
3.1.2 Analisi fattoriale.....	25
3.1.3 Indici di coerenza interna	28
3.1.4 Procedura split-half sample	31
3.2 Verifica della validità del criterio.....	33
3.3 Validazione delle due sottoscale: <i>Aspetti Organizzativi e Azione Didattica</i>	35
3.3.1 Aspetti Organizzativi.....	35
3.3.2 Azione didattica.....	36

3.4	Validazione della scala a 4 passi	38
3.4.1	Statistiche descrittive.....	38
3.4.2	Verifica dell'affidabilità interna	39
3.4.3	Verifica della validità del criterio	45
CAPITOLO 4.....		47
I MODELLI A CLASSI LATENTI		47
4.1	I modelli a classi latenti tradizionali.....	48
4.1.1	Stima dei parametri del modello.....	50
4.1.2	Valutazione dell'adattamento delle stime del modello.....	54
4.1.3	La significatività degli effetti	56
4.1.4	Classificazione.....	57
4.2	I modelli a classi latenti non tradizionali.....	58
4.2.1	Modelli a classi latenti fattoriali	59
4.3	I modelli di regressione a classi latenti.....	61
CAPITOLO 5.....		64
VALIDAZIONE DEL QUESTIONARIO CON L'UTILIZZO DI MODELLI A CLASSI LATENTI.....		64
5.1	Dimensionalità della scala.....	65
5.1.1	Determinazione del numero di fattori.....	65
5.1.2	Descrizione dei livelli identificati.....	68
5.2	Validità del criterio.....	75
5.2.1	Determinazione del numero di cluster.....	75
5.2.1.1	<i>Primo fattore</i>	75
5.2.1.2	<i>Secondo fattore</i>	77
5.2.1.3	<i>Terzo fattore</i>	77
5.2.2	Descrizione dei cluster identificati	78
5.2.3	Verifica della validità del criterio utilizzando le variabili latenti	80
5.2.4	Validità di criterio dell'intera scala	81
CONCLUSIONI		83
BIBLIOGRAFIA		85

INTRODUZIONE

Per *scala di misura* si intende un insieme di item che ha lo scopo di rilevare una variabile non direttamente osservabile, chiamata *variabile latente*. Si sviluppano scale di misura quando si vuole misurare un fenomeno di cui si conosce l'esistenza, ma che non si può valutare in maniera diretta. La scala di misura, dunque, costituisce lo strumento di cui il ricercatore si serve per poter conoscere il costrutto di interesse e in quanto tale è importante assicurarsi che sia robusto sotto il profilo psicometrico. A tale scopo si seguono delle procedure che sono state delineate a partire dalla fine degli anni settanta e che riguardano principalmente la definizione del costrutto d'interesse, la definizione del dominio, l'affidabilità, la validità e la dimensionalità della scala di misura.

Le tecniche statistiche che generalmente si utilizzano nel processo di sviluppo e validazione di una scala si riferiscono a variabili di natura metrica (indici di correlazione, analisi fattoriale,...). Le variabili generate dagli item di un questionario, invece, raramente sono di tipo quantitativo, esse sono spesso variabili ordinali o nominali, che richiedono l'utilizzo di metodi statistici adatti.

Questa tesi si propone di confrontare le tecniche statistiche tradizionali, utilizzate per la validazione di una scala, con un approccio alternativo, che, oltre ad essere più adatto a variabili di tipo categoriale (ordinali o nominali), considera esplicitamente la natura non osservabile del fenomeno oggetto di studio. Questo metodo alternativo consiste nell'*analisi a classi latenti*: con modelli a classi latenti appropriati è possibile verificare le proprietà di affidabilità e validità di una scala di misura.

L'oggetto di riferimento di questa tesi è il *questionario di valutazione della didattica* da parte degli studenti dell'Università degli Studi di Padova. I dati utilizzati si riferiscono all'Anno Accademico 2012/2013 e sono stati raccolti tramite un'indagine via web.

L'organizzazione della tesi è la seguente. Nel Capitolo 1 si introduce una procedura tipica della letteratura psicometrica, illustrata da De Vellis (1991) nella sua opera "*Scale Development. Theory and applications*", utilizzata per lo sviluppo di una scala di misura e articolata in otto fasi: determinazione del costrutto, generazione degli item, determinazione della struttura, revisione degli item, introduzione di item di validazione, somministrazione ad un campione, valutazione degli item e ottimizzazione della lunghezza della scala.

In seguito ci si sofferma su due aspetti chiave per la validazione di un questionario: l'*Affidabilità* e la *Validità*, che a sua volta si divide in *validità di contenuto*, *di criterio* e *di costruito*.

Nel Capitolo 2, viene presentato il progetto PRODID (Preparazione alla Professionalità Docente e Innovazione Didattica) e il questionario a 10 passi, composto principalmente da 18 domande, utilizzato come strumento per valutare la “qualità percepita” dagli studenti dell’Università degli Studi di Padova nei confronti delle attività didattiche. Vengono, dunque, presentati i 18 item sottoposti agli studenti e i dati raccolti dall’indagine.

In seguito, nel Capitolo 3, vengono esposti i risultati ottenuti dalla validazione del questionario tramite l’utilizzo di tecniche statistiche tradizionali, che, come visto sopra, sono più adatte a variabili di natura metrica. In particolare si conduce la verifica dell’affidabilità interna, utilizzando indici di correlazione, analisi fattoriale, indici di coerenza interna e test sulle differenze, e la verifica della validità di criterio, con l’utilizzo di indici di correlazione e dell’ANOVA. Sempre in questo capitolo viene verificata l’affidabilità e la dimensionalità delle due sottoscale Aspetti Organizzativi e Azione Didattica, composte rispettivamente da 4 e 3 item, le quali forniscono le informazioni che sono poi state messe a disposizione di ciascun docente. Nella parte finale del capitolo il processo di validazione viene ripetuto sulla scala trasformata: le 10 categorie di risposta vengono raggruppate in 4. Si effettua, dunque, un confronto tra la scala a 10 passi originale e quella a 4 passi trasformata.

Nel Capitolo 4 vengono introdotti tre diversi tipi di modelli a classi latenti: i modelli a classi latenti tradizionali (di tipo cluster), i modelli fattoriali a classi latenti e i modelli di regressione a classi latenti. Di questi modelli si descrivono principalmente le caratteristiche e i metodi per valutarne la bontà di adattamento.

Infine, nel Capitolo 5, la validazione dello strumento di valutazione della didattica viene effettuata attraverso l’analisi a classi latenti, la quale considera le variabili come ordinali e tiene conto della non osservabilità del costruito d’interesse. Il nuovo approccio viene, dunque, confrontato con quello tradizionale. Vengono, inoltre, fatte alcune considerazioni sul possibile sviluppo di una scala di misura diversa da quella originale a 10 passi.

Nelle conclusioni vengono commentati i risultati ottenuti tramite l’utilizzo dei due approcci, descrivendone le differenze principali e indicando quale dei due risulta preferibile.

CAPITOLO 1

SVILUPPO E VALIDAZIONE DI UNA SCALA DI MISURA: APPROCCIO TRADIZIONALE

Una variabile latente può essere definita come un fenomeno non osservabile. Essa è caratterizzata da due aspetti: non si manifesta in maniera diretta e non è costante, ovvero la sua forza o intensità è soggetta a variazione. Nonostante sia impossibile quantificare in maniera diretta una variabile latente, si può utilizzare una *scala di misura* per valutarne l'intensità. Lo sviluppo di una scala di misurazione mira a stimare la forza di una variabile non osservabile sotto determinate condizioni di spazio e tempo.

In questo capitolo si studia il processo di sviluppo di una scala di misura; in particolare nel primo paragrafo viene illustrato l'approccio affrontato da De Vellis (1991), in cui si analizza la creazione di una scala multi-item e la sua validazione. Nel secondo e terzo paragrafo si approfondiscono due dei concetti chiave per la validazione di un questionario: l'*Affidabilità* e la *Validità*, rispettivamente.

1.1 Processo di sviluppo di una scala di misura

Nella sua opera "*Scale development. Theory and applications*" (1991) De Vellis illustra una procedura per lo sviluppo di una scala di misura per rilevare giudizi di soddisfazione. L'approccio utilizzato si articola in otto fasi. Le prime cinque fasi mirano a costruire il questionario: poste le basi tramite la *definizione del costrutto e del dominio* di interesse si passa alla *generazione di un insieme di item* a cui viene data una *struttura* adeguata, viene poi chiesto ad un *gruppo di esperti* di *revisionare gli item* selezionati e, infine, viene considerata l'introduzione di *item di validazione*. Le restanti tre fasi servono ad ultimare la scala di misura, assicurandosi che essa goda della proprietà di affidabilità: con i dati generati dalla *somministrazione del questionario ad un campione di soggetti* si effettua la *valutazione dell'affidabilità e l'ottimizzazione della scala*, tramite l'utilizzo di analisi statistiche.

Di seguito vengono descritte le otto fasi in cui la procedura è articolata.

1) **Determinare chiaramente ciò che si vuole misurare**

Spesso i questionari vengono utilizzati per indagare fenomeni non osservabili, come, ad esempio, atteggiamenti, comportamenti o preferenze. Questi concetti astratti sono caratterizzati da molti aspetti e difficilmente misurabili in maniera diretta. Il punto di partenza fondamentale nello sviluppo di una scala di misura consiste, dunque, nel fornire una definizione chiara del *costrutto* di interesse. Risulta importante valutare quali aspetti del fenomeno si vogliano enfatizzare e su quali non si desideri porre particolare interesse, definendone il *dominio*. Si può rivelare di grande aiuto lo studio della teoria da cui il fenomeno proviene, per risalire ad una definizione ben formulata del costrutto e ricercare i fenomeni o gli aspetti a cui esso potrebbe essere correlato.

2) **Generare un insieme di item**

Dopo aver definito il costrutto di interesse e il suo dominio si passa alla seconda fase, che consiste nel *generare un insieme di item* in grado di catturare il fenomeno specificato.

In primo luogo si vuole sottolineare il fatto che lo sviluppo di una scala multi-item risulta più ragionevole rispetto alla scelta di un singolo item. La risposta ad un solo quesito difficilmente fornisce una misura esatta della variabile latente sottostante, infatti, in genere il singolo item tende ad avere una bassa correlazione con il costrutto di interesse. Churchill (1979) sottolinea questo aspetto, proprio delle scale di misura basate su un singolo item, riportando un esempio tratto da Jacoby (1978): è piuttosto irragionevole misurare l'intelligenza di un gruppo di individui rivolgendolo solo un quesito; la singola domanda, infatti, non è in grado di catturare tutti gli aspetti coinvolti nella definizione del livello d'intelligenza di un soggetto.

Perciò è importante formulare un numero adeguato di item in grado di catturare diversi aspetti dello stesso fenomeno. Ogni item della scala deve, quindi, riflettere il costrutto d'interesse, "muoversi" nella stessa direzione e con la stessa intensità della variabile latente sottostante.

Nel caso di scala multidimensionale i costrutti di interesse sono di numerosità maggiore di uno: non tutti gli item sono correlati alla stessa variabile non osservabile, ma ogni item risulta collegato ad uno degli aspetti sottostanti.

L'intento è, dunque, quello di catturare il fenomeno d'interesse sviluppando un insieme di item che siano in grado di valutare aspetti differenti dello stesso costrutto.

3) Determinare la struttura della scala di misura

In questa fase viene scelta la struttura con cui presentare gli item. Tra le forme più diffuse di scala di misura ci sono la *scala di Likert* e la *scala del differenziale semantico*.

Nella *scala di Likert* gli item vengono presentati sotto forma di affermazioni e agli intervistati viene chiesto di esprimere il proprio grado di disaccordo o accordo nei confronti di esse. Le risposte possono essere formate da un numero dispari o pari di opzioni, di conseguenza può essere presente o meno una risposta neutrale. I punteggi di ciascun item vengono poi sommati e formano il punteggio totale della scala per ciascun intervistato.

La *scala del differenziale semantico*, invece, prevede che i soggetti intervistati possano valutare il fenomeno di interesse su una scala caratterizzata ai due estremi da due aggettivi, i quali esprimono concetti opposti (caso bipolare) o la presenza/assenza di un certo attributo (caso unipolare).

Entrambe le scale vengono trattate come scale di misura per intervallo, le cui proprietà sono ordine e distanza: le risposte indicano in che misura (maggiore o minore) l'oggetto di interesse viene valutato e i punteggi adiacenti risultano equidistanti.

4) Far revisionare l'insieme di item da un gruppo di esperti

Nella quarta fase, relativa alla *validità del contenuto*, l'insieme di item viene revisionato da un gruppo di persone ben informate sul fenomeno preso in considerazione. La revisione condotta da persone esperte può, infatti, confermare o invalidare la definizione del costrutto di interesse. Gli esperti sono in grado di valutare se gli item catturano realmente il costrutto (o i costrutti) e analizzare la chiarezza nell'esposizione dei quesiti. In questa fase vengono eliminati gli item ritenuti inadeguati o non rilevanti, riformulati gli item poco chiari e inclusi aspetti del fenomeno precedentemente ignorati.

5) Considerare l'introduzione di item di validazione

In alcuni casi può risultare conveniente aggiungere all'insieme degli item principali altri quesiti che aiutino nella determinazione della validità della scala. Esistono due tipologie di "*validation items*".

Il primo tipo serve ad eliminare difetti o problemi nel questionario. All'inizio del questionario possono essere poste alcune domande in grado di valutare se esistano altre motivazioni, oltre all'oggetto di interesse, capaci di influenzare le risposte degli intervistati. Per esempio se un individuo desidera presentarsi alla società in un modo considerato positivo, allora le sue risposte agli item potrebbero essere distorte; inserendo una scala che misuri la "desiderabilità sociale" è possibile valutare in anticipo quanto le risposte agli item siano influenzate da questa variabile e, quindi, quanto siano soggette a distorsione (De Vellis, 1991).

La seconda classe include tutti gli item necessari a valutare la validità del costrutto. Come si vedrà più avanti, infatti, la verifica della validità del costrutto consiste nell'analizzare la coerenza dell'oggetto misurato con altri fenomeni ad esso teoricamente collegati. Per poter condurre questa analisi è quindi necessario somministrare ai rispondenti dei quesiti in grado di misurare i possibili fenomeni correlati.

6) Somministrare gli item ad un campione di soggetti

A questo punto il questionario risulta pronto per essere somministrato ad un gruppo di rispondenti. Il campione deve essere preferibilmente numeroso, ma non esiste un preciso criterio sulla sua dimensione. Nunnally (1978), per esempio, suggerisce una numerosità pari a 300.

7) Valutare gli item

La settima fase è considerata la parte principale nel processo di sviluppo di una scala di misura e consiste nel valutare l'*affidabilità della scala*.

Supponendo che gli item che compongono una scala siano tutti fortemente legati al costrutto sottostante, allora dovrebbero anche risultare strettamente collegati tra loro, segnalando un alto grado di coerenza interna. La prima caratteristica che si desidera in un insieme di item è, quindi, un livello di *inter-correlazione* elevato. Uno strumento utile per questa verifica è la matrice delle correlazioni: elevati valori di correlazione tra item indicano elevati livelli di affidabilità.

Un importante indice generale per la verifica della coerenza interna è il coefficiente *Alpha di Cronbach*. Questo indicatore determina la proporzione di varianza condivisa dagli item che è attribuibile al costrutto indagato e assume valori compresi tra 0 e 1:

valori vicini allo zero indicano un basso grado di affidabilità tra item, mentre valori prossimi a 1 indicano un alto grado di coerenza della scala. Nunnally (1978) suggerisce come soglia per l'Alpha il valore 0,70: al di sotto di questo valore il grado di affidabilità risulta basso. Mentre De Vellis (1991) specifica sei classi di valori: sotto lo 0,60 valori inaccettabili, tra 0,60 e 0,65 valori indesiderabili, tra 0,65 e 0,70 valori minimamente accettabili, tra 0,70 e 0,80 valori rispettabili, tra 0,80 e 0,90 valori molto buoni, infine, nel caso di valori superiori allo 0,90 la scala risulta molto affidabile e perciò si può valutare l'ottimizzazione della sua lunghezza per garantire una compilazione più veloce e meno "faticosa" ai soggetti intervistati.

8) Ottimizzare la lunghezza della scala

In quest'ultima fase del processo si suppone di essere in possesso di uno strumento di misurazione in cui tutti gli item risultano coerenti e formano una scala affidabile (con livello di Alpha superiore a 0,90).

Generalmente i rispondenti preferiscono compilare questionari più corti, perciò in questa fase è importante considerare l'*ottimizzazione della lunghezza della scala*. Considerando, però, che scale più lunghe tendono ad essere più affidabili, ci si pone di fronte ad un problema di trade-off tra brevità e affidabilità.

Si potrebbe, per esempio, considerare l'eliminazione di quegli item che contribuiscono in maniera minore alla coerenza interna della scala. Uno strumento utile per effettuare questa verifica è l'Alpha di Cronbach calcolato escludendo l'item che si vuole analizzare. Se il valore di questo indice supera il valore dell'Alpha calcolato sulla scala nel suo complesso, allora, eliminando quell'item, l'affidabilità complessiva della scala aumenterebbe. Questo strumento risulta, dunque, di supporto alla decisione di accorciare la scala di misura.

1.2 Affidabilità

In questo paragrafo si esamina l'affidabilità quale proprietà fondamentale per una scala di misura e si analizza la logica sottostante gli strumenti di verifica utilizzati nella validazione del questionario di soddisfazione della didattica da parte degli studenti, oggetto di questa tesi.

L'affidabilità è una condizione necessaria, ma non sufficiente, affinché una scala di misura risulti valida. Peter (1979) definisce l'affidabilità come il grado in cui una misura è libera da errore e, quindi, produca risultati coerenti.

Considerando la seguente relazione:

$$X_O = X_T + X_S + X_R^1 \quad (1)$$

dove:

- X_O = punteggio osservato;
- X_T = vero punteggio della variabile latente (non osservabile);
- X_S = errore sistematico (i.e. caratteristiche dell'intervistato che influenzano le risposte);
- X_R = errore casuale (i.e. umore degli intervistati, livello di stanchezza);

si può affermare che una scala di misura è perfettamente affidabile quando $X_R = 0$, ovvero quando la variazione del valore osservato dal vero punteggio non è dovuta ad errori casuali.

1.2.1 Indici di correlazione

Come già accennato nel punto 7) del Paragrafo 1.1, una scala è affidabile se gli item che la compongono sono correlati tra loro. Per verificare ciò si utilizza il *coefficiente di correlazione di Pearson*, che consente di verificare il grado di associazione tra i vari item presi due a due (*inter-item correlation*) oppure il grado di correlazione tra item e la scala nel suo complesso (*item-to-total correlation*). Esistono due tipi di *item-to-total correlation*: nel primo tipo il singolo item viene messo in relazione al totale della scala escluso l'item stesso (*corrected item-to-total correlation*); nel secondo l'item in

¹ Churchill (1979)

questione viene valutato assieme alla scala nel suo totale, item compreso (*uncorrected item-to-total correlation*). In generale si utilizza la versione corretta; un item che presenta un valore elevato di questo indice risulta essere molto coerente al resto della scala, viceversa, un item con un indice di correlazione item-to-total basso è poco coerente rispetto al resto della scala. Item con bassa coerenza devono essere studiati tramite ulteriori analisi statistiche (Analisi Fattoriale, Alpha di Cronbach, ecc), affinché venga valutata la loro possibile esclusione dal questionario. Hair, Black, Babin, Anderson & Tatham (2006) suggeriscono come soglia minima per le correlazioni tra item il valore 0,30, mentre le correlazioni item-to-total dovrebbero superare lo 0,50: valori che non superano queste soglie indicano item non coerenti con la scala di misura, la quale, di conseguenza, risulta non affidabile.

1.2.2 Analisi fattoriale

Nella fase di verifica dell'affidabilità di una scala di misura è importante verificarne anche la dimensionalità, ovvero la presenza di una o più variabili latenti sottostanti. A questo scopo si utilizza l'analisi fattoriale, ovvero una tecnica statistica che consente di ridurre il numero di variabili analizzate tramite l'estrazione di fattori, capaci di spiegare aspetti non osservabili. Questa tecnica risulta utile per due motivi fondamentali: per prima cosa consente di determinare quante variabili latenti sono alla base della scala di misura, e in secondo luogo permette di utilizzare un numero ridotto di fattori per spiegare la variazione tra un gran numero di variabili originali. Essa, inoltre, permette di fornire un significato ad ogni fattore, identificando gruppi di item che sono collegati allo stesso fattore latente.

Per l'estrazione dei fattori (o componenti) si utilizza generalmente il *Metodo delle componenti principali*, il quale considera i fattori come una combinazione lineare ponderata delle variabili originali. Le componenti assumono, dunque, questa espressione:

$$\begin{aligned} f_1 &= w_{11}y_1 + w_{12}y_2 + \dots + w_{1p}y_p \\ &\dots \\ f_p &= w_{p1}y_1 + w_{p2}y_2 + \dots + w_{pp}y_p \end{aligned}$$

dove:

- p = numero variabili originali;
- y_1, \dots, y_p = variabili originali (item del questionario);
- f_1, \dots, f_p = fattori (o componenti);
- w_{ki} = coefficienti fattoriali, con $k = 1, \dots, p$ e $i = 1, \dots, p$ (k indica il fattore e i la variabile originale).

Per la scelta del numero di fattori da selezionare si utilizza comunemente la *Regola degli autovalori maggiori di 1*, secondo la quale si dovrebbero prendere in considerazione solo quelle componenti che hanno un autovalore maggiore o uguale ad uno.

Infine, per semplificare l'interpretazione delle componenti estratte si utilizzano dei metodi di rotazione dei fattori che permettono di ottenere soluzioni più chiare ed evidenti. Uno tra i più comuni metodi di rotazione ortogonale è chiamato *Varimax*, che massimizza la somma delle varianze dei quadrati dei coefficienti fattoriali, amplificando i coefficienti più alti, riducendo quelli più bassi e mantenendo l'incorrelazione tra fattori.

1.2.3 Alpha di Cronbach

Come già visto nel Paragrafo 1.1, per verificare l'affidabilità di una scala si analizza in particolare la sua coerenza interna. Gli item che compongono una scala dovrebbero essere logicamente collegati alla variabile latente sottostante, se questa relazione risulta forte ne consegue che gli item sono molto correlati tra loro.

Tra gli strumenti di verifica della coerenza interna, oltre l'indice di correlazione tra item e quello tra item e totale della scala, si è accennato al coefficiente *Alpha di Cronbach*. Per comprendere il calcolo dell'Alpha, si decide di denominare la variabile latente sottostante la scala di misura X . Gli item che compongono una scala condividono la varianza causata da X . Al variare di X , dunque, variano anche gli item ad esso collegati. La variazione di ciascun item è però dovuta non solo alla variabile latente, ma anche ad un termine d'errore. Questa fonte di errore non è condivisa, ma è propria di ogni item. La varianza totale di ogni item è, dunque, una combinazione tra la varianza condivisa, dovuta alla variabile latente X , e la varianza "unica" dovuta al termine d'errore. Studiando la definizione di De Vellis (1991) che descrive l'affidabilità come la proporzione di varianza attribuibile al *vero punteggio* (*true*

score) della variabile latente, si giunge alla determinazione di Alpha come funzione del rapporto tra varianza comune (o condivisa) e varianza totale.

Si consideri la matrice di varianza e covarianza (di dimensione p) di una scala di misura, in cui sulla diagonale principale sono presenti le varianze di ogni item (σ_i^2) e fuori dalla diagonale sono presenti le covarianze tra item (σ_{ij} con $i \neq j$). Le varianze σ_i^2 rappresentano variazioni che ciascun item condivide con se stesso, si possono dunque definire uniche per ogni item. Al contrario i termini σ_{ij} rappresentano variazioni condivise tra item. La somma di tutti gli elementi presenti nella matrice rappresenta la varianza totale. Dunque, la proporzione di varianza unica o non condivisa può essere espressa nel modo seguente:

$$\sum_{i=1}^p \sigma_i^2 / \sigma_{totale}^2$$

dove:

- p = numero di item che compongono la scala;
- σ_i^2 = varianza del singolo item ($i=1, \dots, p$);
- $\sigma_{totale}^2 = \sum_{i=1}^p \sigma_i^2 + 2 \sum_i^p \sum_j^p \sigma_{ij}$ con $i > j$ (somma di tutti i p^2 elementi presenti nella matrice di varianza e covarianza).

Di conseguenza la proporzione di varianza condivisa può essere espressa come:

$$1 - \left(\sum_{i=1}^p \sigma_i^2 / \sigma_{totale}^2 \right)$$

Moltiplicando questa espressione per un fattore di correzione, il coefficiente Alpha di Cronbach risulta pari a:

$$\alpha = \frac{p}{p-1} \left(1 - \frac{\sum_{i=1}^p \sigma_i^2}{\sigma_{totale}^2} \right)$$

1.2.4 Indici di coerenza interna basati sul metodo *split-half*

Oltre al coefficiente Alpha esistono altri indici in grado di misurare la coerenza interna di una scala. Essi si basano sulla tecnica dello *split-half items*, ovvero la suddivisione della scala in due gruppi omogenei di item. Se una scala risulta affidabile, insiemi di item ricavati da essa dovrebbero fornire risultati simili in termini di coerenza interna.

La suddivisione può essere condotta con vari metodi. Il primo consiste nel separare la prima metà di item dalla seconda; questo tipo di suddivisione risulta problematico perché potrebbero esistere dei fattori che influenzano i due gruppi di item in maniera differente: per esempio gli intervistati potrebbero rispondere alla parte finale di un questionario con meno attenzione rispetto alla parte iniziale, a causa del loro livello di “stanchezza”; questa componente di errore porterebbe, dunque, ad una differenziazione tra i due gruppi che non è causata dalla differenza degli item stessi, ma da una caratteristica o da uno stato d’animo dei rispondenti.

Per evitare questo problema conviene suddividere la scala in item codificati con numero pari e item codificati con numero dispari.

In particolare si calcolano tre indici: l’indice *split-half R*, che coincide con il coefficiente di correlazione tra i punteggi totali dei due sottogruppi di item e quindi varia tra -1 e +1; l’indice *Y di Spearman-Brown* che è pari a $2|R|/(1 + |R|)$, che varia tra 0 e 1; infine il coefficiente *G di Guttman* pari a $2 \left(1 - \frac{\sigma_{S1}^2 + \sigma_{S2}^2}{\sigma_S^2}\right)^2$ (dove σ_{S1}^2 e σ_{S2}^2 rappresentano le varianze dei due sottogruppi di item S1 e S2 ; σ_S^2 è la varianza complessiva). Valori di questi indici prossimi ad 1 indicano alti livelli di affidabilità.

1.2.5 Procedura *split-half sample*

Un’ulteriore procedura per verificare l’affidabilità interna di una scala viene chiamata *split-half sample* e consiste nel suddividere il campione complessivo in due sottocampioni omogenei; si sottolinea l’importanza di considerare due sottogruppi estratti casualmente dal campione complessivo in modo da permettere un confronto tra essi. A questi due sottogruppi di soggetti si applica il calcolo degli indici di coerenza interna illustrati nei Sottoparagrafi 1.2.3 e 1.2.4. Se il valore degli indici risulta simile nei due sottocampioni, allora si può

² Guido, Bassi, Peluso (2010)

affermare che la scala è affidabile in quanto è in grado di misurare lo stesso costrutto su due gruppi differenti di soggetti. Tramite questa tecnica è, inoltre, possibile condurre a livello di singolo item un *test sulla differenza delle medie* nei due sottocampioni. Utilizzando il t-test, infatti, si è in grado di verificare se la differenza tra i punteggi medi di ciascun item è significativa o meno. Risultati non significativi indicano un alto grado di coerenza interna.

1.3 Validità

L'affidabilità riguarda quanto una variabile non osservabile influenzi un insieme di item, mentre la validità è la proprietà della scala che si riferisce al grado in cui il costrutto latente è la causa della correlazione tra item. Una scala di misura è valida se l'oggetto da essa misurato coincide con il fenomeno di interesse. In termini analitici ciò si traduce in $X_O = X$, ovvero il punteggio osservato deve essere uguale al vero punteggio della variabile latente. Osservando l'equazione (1) i due punteggi risultano uguali se $X_S = X_R = 0$ (errore sistematico e errore casuale uguali a zero), confermando il fatto che l'affidabilità di una scala non ne garantisce la validità, ma è solo una condizione necessaria.

Esistono tre tipi di validità:

- Validità di contenuto
- Validità di criterio
- Validità di costrutto.

1.3.1 Validità di contenuto

La validità di contenuto si riferisce al grado in cui un insieme di item riflette il dominio di interesse. Al fine di verificare questa proprietà ci si pone la domanda "In che misura la scala sviluppata valuta effettivamente il fenomeno di interesse?" (Ghiselli, Campbell, Zedeck, 1981). Per essere valida una scala non solo deve indagare il costrutto di interesse, ma deve riuscire a ricoprire l'intero dominio, rappresentandone tutti gli aspetti. Perciò la validità del contenuto risulta più semplice quando il dominio è ben definito. Il metodo di verifica della validità del contenuto consiste nel sottoporre ad esame l'insieme di item: si seleziona un

gruppo di esperti che possano dare giudizi sull'adeguatezza della scala e possano suggerire eventuali aspetti non inclusi.

1.3.2 Validità di criterio

La validità di criterio si riferisce al grado in cui il punteggio di una scala sia associato ad un'altra variabile, chiamata criterio o *golden standard*. La variabile-criterio deve essere preventivamente inserita nel questionario e consiste, generalmente, in una domanda volta a valutare il costrutto in maniera diretta. Per godere della proprietà della validità del criterio una scala deve variare nella stessa direzione e con la stessa intensità del golden standard.

Per verificare se le due misure variano nella stessa direzione si calcola il coefficiente di correlazione tra il punteggio complessivo della scala e quello della variabile-criterio. Elevati valori dell'indice di correlazione indicano un alto grado di validità di criterio.

Per verificare se esse variano con la medesima intensità si utilizza *l'Analisi della Varianza (ANOVA) univariata*. Questa analisi è utile per valutare se a livelli crescenti di una variabile corrispondono analoghi incrementi dell'altra, in quanto verifica se le differenze tra i punteggi medi della scala, considerata come variabile dipendente, assunti in corrispondenza dei diversi livelli della variabile-criterio, considerata come variabile indipendente, sono statisticamente significative. Differenze significative sono segnale di un buon grado di validità della scala.

1.3.3 Validità di costrutto

La validità del costrutto si riferisce al grado in cui il costrutto misurato dalla scala risulta coerente con misure di fenomeni teoricamente collegati. Per misurare questi fenomeni vengono inseriti nel questionario dei quesiti (*validation items*) che, secondo la letteratura di riferimento, possono essere cause o conseguenze del costrutto indagato. Anche in questo caso le analisi da condurre sono due: il calcolo dei coefficienti di correlazione tra punteggio complessivo della scala e punteggio assunto dai validation item e l'Analisi della Varianza per verificare se le singole misure dei fenomeni associati variano in maniera coerente ai diversi livelli del costrutto indagato.

CAPITOLO 2

PROGETTO PRODID E QUESTIONARIO RIVOLTO AGLI STUDENTI DELL'UNIVERSITA' DEGLI STUDI DI PADOVA

2.1 Il progetto PRODID

PRODID (Preparazione alla Professionalità Docente e Innovazione Didattica) è un progetto approvato dal Consiglio di Amministrazione dell'Ateneo di Padova nel luglio 2013, che si pone l'obiettivo di costituire e sviluppare strategie di sostegno alla professionalità docente, nella logica di una crescita progressiva delle competenze per la didattica e l'insegnamento. In linea con le più accreditate prospettive internazionali, si propone di costituire un Teaching and Learning Center (TLC), in grado di coinvolgere attivamente i docenti verso un innalzamento continuo della capacità professionale.³

Nel progetto sono coinvolte quattro Unità di Ricerca (UR):⁴

- UR1, *Metodologie di formazione per l'insegnamento-apprendimento*: è un'area chiamata a predisporre i percorsi di formazione dei soggetti coinvolti riguardo alla didattica facendo riferimento alla progettazione, realizzazione e valutazione di interventi formativi e alla qualità delle dinamiche fra chi insegna e chi apprende.
- UR2, *Innovazione didattica e tecnologie*: ha il compito di concretizzare la messa a punto dei processi di innovazione della didattica con particolare riferimento all'impiego di tecnologie per la didattica e dell'uso di modalità di intervento on line.
- UR3, *Ricerca valutativa nel campo della didattica*: è focalizzata sulla ricerca in ambito valutativo e particolarmente attenta a garantire e a promuovere adeguati livelli di approfondimento scientifico delle pratiche di valutazione realizzate in ateneo e ad un proficuo utilizzo dei risultati nei contesti di insegnamento.

³ <http://www.unipd.it/progetto-prodid>

⁴ <https://elearning.unipd.it/prodid/mod/page/view.php?id=28>

- UR4, *Modelli organizzativo-gestionali per la qualità della didattica*: è orientata a predisporre uno studio del contesto in cui si realizzano efficaci interventi didattici, sulla base di una analisi del sistema e del campo organizzativo-strutturale che supportano la didattica e la formazione erogata.

2.2 Il questionario

Il riferimento principale del lavoro di analisi e validazione svolto dall'Unità di Ricerca n.3 è il questionario rivolto agli studenti via web nell'Anno Accademico 2012/2013. Il questionario ha lo scopo di valutare la "qualità percepita" dagli studenti sull'attività didattica svolta nel corso dell'anno. La misurazione della soddisfazione da parte degli studenti permette di fornire informazioni sull'andamento del corso al fine di supportare il processo di miglioramento della didattica in Ateneo.

La prima raccolta delle opinioni espresse dagli studenti dell'Ateneo di Padova risale al primo semestre dell'Anno Accademico 1999/2000, quando tutte le facoltà dell'Ateneo aderirono all'indagine statistica (in precedenza c'erano state numerose sperimentazioni, che avevano però coinvolto solo alcune Facoltà). Fino all'a.a. 2009/2010 il questionario è stato somministrato agli studenti sotto forma di foglio cartaceo: durante le lezioni dei rilevatori appositamente formati distribuivano i questionari, i quali venivano compilati dagli studenti presenti al momento della rilevazione. Dall'a.a. 2010/2011 è stato implementato il modello di indagine online di tipo CAWI (Computer Assisted Web Interview), attraverso un sistema di questionari sottoposti agli studenti tramite Uniweb, il sistema informativo dell'Università degli Studi di Padova. Allo studente vengono proposti i questionari di valutazione per le attività (di almeno 15 ore) che nel libretto online appartengono all'Anno Accademico di offerta in corso.⁵

Il questionario per la valutazione della didattica sottoposto agli studenti dell'Ateneo di Padova nell'anno accademico 2012/2013 è composto da due domande filtro iniziali, una sulla volontà di compilazione (se la risposta è negativa il questionario è terminato) e l'altra sulla percentuale di frequenza alle lezioni (la modalità "meno del 30% delle lezioni frequentate")

⁵ Technical Report Series, N.1, "Tools for teaching assessment at Padua University: role, development and validation. PRODID project (Teacher professional development and academic educational innovation) – Report of the Research Unit n.3; May 2015.

identifica i non frequentanti, ai quali viene sottoposta solo una parte del questionario più una domanda sul motivo della mancata frequenza).

Il questionario completo rivolto agli studenti che hanno frequentato il corso è composto da 18 quesiti, più 3 rivolti solo agli studenti delle Lauree Magistrali. Gli studenti hanno risposto a tutti i quesiti assegnando un punteggio da 1 a 10 a ciascuno, dove 10 corrisponde alla situazione ottimale.

Di seguito viene presentato il questionario completo.

1. D01. *All'inizio del corso gli obiettivi e i contenuti sono stati definiti in modo chiaro?*
2. D02. *Le modalità d'esame sono state definite in modo chiaro?*
3. D03. *Gli orari delle lezioni sono stati rispettati?*
4. D04. *Le ore previste sono in numero adeguato per lo svolgimento del programma?*
5. D05. *Le conoscenze preliminari sono sufficienti per la comprensione degli argomenti?*
6. D06. *Il docente stimola/motiva l'interesse verso la disciplina?*
7. D07. *Il docente espone gli argomenti in modo chiaro?*
8. D08. *Il materiale didattico consigliato è adeguato?*
9. D09. *Il docente è disponibile nei confronti delle esigenze degli studenti?*
10. D10. *Il docente è stato reperibile durante gli orari di ricevimento?*
11. D11. *Esercitazioni/laboratori/seminari, se previsti, sono adeguati?*
12. D12. *Le aule in cui si svolgono le lezioni sono adeguate?*
13. D13. *I locali e le attrezzature per i laboratori sono adeguati?*
14. D14. *Complessivamente quanto è soddisfatto del corso?*
15. D15. *Il carico di studio richiesto è equilibrato rispetto al numero di CFU assegnati?*
16. D16. *Indipendentemente da come si è svolto il corso, quanto è interessato ai contenuti?*
17. D17. *Quanto ritiene coerente l'insegnamento rispetto agli obiettivi del corso di studi?*
18. D18. *Ritiene che l'insegnamento fornisca competenze adeguate in ambito lavorativo?*

Agli studenti delle Lauree Magistrali, inoltre, vengono poste altre tre domande.

Considerando il percorso di LT, quanto i contenuti sono:

LM01. Ripetizione di quanto già studiato

LM02. Approfondimento di quanto già affrontato

LM03. Argomenti del tutto nuovi

Le informazioni messe a disposizione di ciascun docente, per ciascun insegnamento da questi tenuto, riguardano 3 indicatori:

- media globale dell'indicatore di **soddisfazione generale (item 14)**;
- indicatore relativo agli **aspetti organizzativi = media aritmetica** dei punteggi attribuiti agli **item 1** (chiarezza sugli obiettivi del corso), **2** (chiarezza sulle modalità d'esame), **3** (rispetto degli orari di lezione) e **8** (adeguatezza del materiale didattico);
- indicatore relativo **all'azione didattica = media aritmetica** dei punteggi attribuiti agli **item 6** (stimolo/motivazione verso la disciplina), **7** (chiarezza espositiva) e **9** (attenzione verso le esigenze degli studenti); dal 2013/2014 l'item 9 è stato escluso dall'indicatore.

2.3 I dati

Il numero di record a disposizione per le analisi effettuata in questa tesi è pari a **163.626**. I questionari raccolti nell'anno accademico 2012/2013 sono **253.318**, dei quali effettivamente compilati **196.103** (77,4% del totale). Da questi dati sono stati eliminati anche **30.197** record relativi agli studenti che hanno frequentato meno del 50% delle lezioni; sono stati, inoltre, eliminati **2.272** record relativi agli studenti iscritti ai corsi singoli e agli studenti Erasmus; infine sono stati esclusi gli **8** questionari compilati per errore.

Il file di dati utilizzato per questa tesi è caratterizzato, dunque, dalle risposte ai questionari relativi a studenti iscritti alle Lauree Triennali, Magistrali o a Ciclo Unico dell'Ateneo di Padova nell'Anno Accademico 2012/2013, che abbiano frequentato almeno il 50% delle lezioni del corso di cui hanno compilato il questionario.

2.3.1 Statistiche descrittive

Dalla Tabella 1 si osserva che gli item 10, 11 e 13 presentano una grande quantità di mancate risposte rispetto agli altri item, indicando che molti studenti hanno preferito non rispondere o rispondere "non so/non pertinente" a questi tre quesiti.

La media delle risposte ai quesiti va dal valore minimo pari a 7,33 in corrispondenza dell'item 15 fino al massimo pari a 8,34 in corrispondenza del quesito 3. In tutti i quesiti, dunque, si

sono riscontrati valori medi alti; ciò indica un livello medio di soddisfazione alto per quanto riguarda tutti gli aspetti che il questionario intende indagare. Si ricorda però che si sono considerati i questionari provenienti da tutti i Corsi di Laurea dell'Università degli Studi di Padova. Potrebbe essere utile e interessante, per lo scopo del progetto di miglioramento della didattica, analizzare come variano i valori medi tra Scuole, all'interno di ogni Scuola, tra Dipartimenti e più dettagliatamente all'interno di ogni Dipartimento, per corso di laurea, tra corsi di Laurea Triennale, Magistrale e a Ciclo Unico.

Anche i valori della deviazione standard sembrano essere stabili tra gli item, variando tra 1,77 e 2,13.

Item	Numero questionari	Media	Deviazione Standard
D01 obiettivi	159.173	7,92	1,82
D02 modalità	158.225	8,00	1,90
D03 orari	160.417	8,34	1,77
D04 numero ore	146.751	7,71	1,97
D05 preconoscenze	160.378	7,35	1,98
D06 stimolo	160.387	7,55	2,13
D07 chiarezza	160.374	7,61	2,09
D08 materiale	159.998	7,49	2,05
D09 disponibile	159.909	8,11	1,86
D10 ricevimento	77.905	8,21	1,87
D11 esercitazioni	98.071	7,75	2,00
D12 aule	160.327	7,52	2,11
D13 attrezzature	100.135	7,53	2,09
D14 soddisfazione complessiva	160.272	7,51	1,96
D15 carico	160.079	7,33	2,09
D16 interesse	160.200	7,99	1,88
D17 coerente	157.504	8,19	1,85
D18 utile	149.151	7,71	2,01

Tabella 1: Statistiche descrittive dei 18 item del questionario.

Come si può notare dalla Tabella 2, la maggior parte degli studenti che ha compilato il questionario proviene da Corsi di Laurea Triennale, a seguire quelli provenienti da Corsi di Laurea a Ciclo Unico e da Corsi di Laurea Magistrale. L'86,9% degli studenti presi in considerazione ha frequentato più del 70% delle lezioni del corso, con percentuali di frequenza simili tra i tre corsi di laurea considerati.

Frequenza alle lezioni	Tipo di corso di laurea a cui è iscritto lo studente							
	Laurea a Ciclo Unico		Laurea Magistrale		Laurea Triennale		Totale	
	Conteggio	%	Conteggio	%	Conteggio	%	Conteggio	%
Tra 50% e 70%	3.462	11,6	3.817	14,1	14.077	13,2	21.356	13,1
Più del 70%	26.467	88,4	23.167	85,9	92.636	86,8	142.270	86,9
Totale	29.929	100	26.984	100	106.713	100	163.626	100

Tabella 2: Numero di questionari per livello di frequenza alle lezioni e tipo di corso di laurea

CAPITOLO 3

VALIDAZIONE DEL QUESTIONARIO

Il controllo della validazione del questionario viene effettuato con lo scopo di misurare l'accuratezza della scala utilizzata e la generalizzabilità dei risultati ottenuti.

Quest'analisi dei dati prevede tre fasi:

- 1) verifica dell'affidabilità interna;
- 2) validità del criterio;
- 3) validità del costrutto.⁶

Le prime due verifiche sono presentate nei paragrafi 3.1 e 3.2 di questo capitolo, invece, non è stato possibile effettuare la verifica della validità del costrutto. Essa, infatti, consiste nell'esaminare se il costrutto indagato, nel nostro caso la soddisfazione degli studenti nei confronti del corso frequentato, è correlato ad aspetti logicamente e/o teoricamente collegati. Nel caso di un questionario di soddisfazione di un consumatore nei confronti di un prodotto acquistato, per esempio, possono essere poste alcune domande riguardanti le fasi di post acquisto (intenzione di riacquisto, passaparola positivo, lamenti) logicamente correlate con il costrutto indagato. Nel caso del questionario preso in considerazione per questo lavoro di tesi non sono presenti, invece, domande che indagano aspetti successivi alla frequentazione del corso, perciò non è stata eseguita la verifica della validità del costrutto.

Le seguenti analisi sono già state effettuate in precedenza da membri dell'Unità di Ricerca N.3, coinvolta nel progetto PRODID, e pubblicate nel Technical Report "*Tools for teaching assessment at Padua University: role, development and validation*".⁷

⁶ Le analisi presentate in questo capitolo sono state effettuate con il software Spss.

⁷ Consultabile al link <http://paduaresearch.cab.unipd.it/8851/>

3.1 Verifica dell'affidabilità interna

La verifica dell'affidabilità interna ha lo scopo di verificare la coerenza tra i diversi item che compongono il questionario, controllando che essi misurino tutti lo stesso costrutto, ovvero la soddisfazione degli studenti nei confronti dei corsi da essi frequentati, e che siano in grado di riprodurre risultati simili con la stessa o un'analogia popolazione.

3.1.1 Analisi di correlazione

L'analisi di correlazione permette di verificare quanto gli item siano correlati col costrutto indagato e tra loro. Il *coefficiente di correlazione di Pearson* (r) consente di verificare il grado di associazione tra ciascun item e la scala nel suo complesso (*item-to-total correlation*) oppure il grado di correlazione tra i vari item presi due a due (*inter-item correlation*). Per il calcolo del primo coefficiente si è proceduto con un'eliminazione listwise, ovvero sono stati esclusi dall'analisi tutti i casi che presentavano una risposta mancante in almeno un item, nel secondo caso è stato possibile effettuare un'eliminazione pairwise in modo da poter utilizzare il maggior numero di dati possibili, ovvero per ogni coppia di item analizzata sono stati eliminati quei casi che presentavano risposta mancante su uno o entrambi gli item.

Dopo aver eseguito l'eliminazione listwise, i record che rimangono inclusi nell'analisi sono 54.963 (il 33,6% del totale) e corrispondono ai record che presentano risposta ad ognuno dei 17 item (viene escluso l'item 14, considerato come golden standard).

La Tabella 3 illustra l'*item-to-total correlation*, ovvero l'indice di correlazione tra singolo item e totale della scala; l'indice è riportato nella forma corretta, in quanto viene calcolato con riferimento al totale della scala escludendo l'item considerato. Si ritengono accettabili valori di questo indice (*coefficiente di Pearson*) maggiori di 0,30. Dalla tabella si osserva che l'indice supera di molto lo 0,30 in corrispondenza di tutti i 17 item; sembra, dunque, esserci un elevato grado di correlazione tra ogni item e la scala nel complesso. Gli item che presentano i valori più bassi sono il numero 5, 12 e 13, relativi a preconcoscenze, adeguatezza di aule e laboratori/attrezzature rispettivamente.

Item	Item-to-total correlation
D01 obiettivi	0,865
D02 modalità	0,832
D03 orari	0,792
D04 numero ore	0,812
D05 preconoscenze	0,718
D06 stimolo	0,877
D07 chiarezza	0,877
D08 materiale	0,855
D09 disponibile	0,862
D10 ricevimento	0,847
D11 esercitazioni	0,850
D12 aule	0,618
D13 attrezzature	0,671
D15 carico	0,783
D16 interesse	0,832
D17 coerente	0,806
D18 utile	0,787

Tabella 3: Correlazione di ciascun item con il totale della scala.

La Tabella 4 rappresenta la matrice di correlazione tra i 18 item ed è stata creata utilizzando l'eliminazione a coppie. I coefficienti di correlazione sono tutti significativi (tutti i p-value sono $< 0,001$) e superano tutti il valore 0,30; si può dunque affermare che sembra esserci una correlazione abbastanza forte tra tutti gli item, i quali dovrebbero indagare aspetti diversi dello stesso costrutto. Si osserva che l'item 14, relativo alla soddisfazione generale, è fortemente correlato a quasi tutti gli altri item; in particolare, si nota una forte correlazione tra questo e gli **item 6** (stimolo) e **7** (chiarezza), mentre le correlazioni minori si presentano in corrispondenza degli **item 12** (aule) e **13** (attrezzature). Questi ultimi item risultano poco correlati con tutti gli altri item, ma molto correlati tra loro (coefficiente di Pearson pari a 0,799 tra item 12 e 13); sembra essere, quindi, confermato il fatto che questi due item indagano un aspetto non direttamente collegato alla soddisfazione della didattica. Si osservano indici molto alti negli incroci tra gli **item 6 e 7**, e tra gli **item 9 e 10**, i quali misurano rispettivamente i primi lo stimolo all'interesse verso la disciplina e la chiarezza nell'esporre gli argomenti e gli ultimi la disponibilità del docente nei confronti delle esigenze degli studenti e la reperibilità del docente durante gli orari di ricevimento.

	D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	D11	D12	D13	D15	D16	D17	D18
D01 obiettivi	1																
D02 modalità	,716	1															
D03 orari	,558	,549	1														
D04 numero ore	,593	,542	,557	1													
D05 preconoscenze	,477	,404	,342	,464	1												
D06 stimolo	,706	,591	,486	,547	,507	1											
D07 chiarezza	,744	,618	,515	,573	,492	,850	1										
D08 materiale	,659	,590	,491	,565	,477	,688	,736	1									
D09 disponibile	,647	,621	,582	,550	,415	,694	,698	,656	1								
D10 ricevimento	,703	,696	,713	,635	,503	,711	,715	,688	,877	1							
D11 esercitazioni	,651	,606	,547	,680	,514	,681	,677	,666	,681	,762	1						
D12 aule	,328	,312	,306	,351	,300	,308	,309	,324	,339	,434	,426	1					
D13 attrezzature	,433	,411	,398	,464	,381	,412	,410	,436	,437	,528	,561	,799	1				
D15 carico	,531	,493	,422	,605	,469	,564	,565	,567	,538	,586	,600	,338	,441	1			
D16 interesse	,580	,489	,423	,445	,470	,696	,627	,566	,548	,630	,587	,298	,400	,521	1		
D17 coerente	,527	,450	,407	,407	,433	,582	,539	,506	,506	,611	,548	,291	,397	,466	,776	1	
D18 utile	,522	,433	,363	,420	,409	,573	,536	,505	,469	,564	,571	,287	,403	,459	,691	,777	1
D14 sodd complessiva	,748	,638	,535	,627	,501	,844	,840	,732	,707	,727	,733	,359	,475	,641	,710	,605	,603

Tabella 4: Inter-item correlations.

3.1.2 Analisi fattoriale

L'*analisi fattoriale* viene utilizzata in questo contesto per indentificare se dai dati raccolti è possibile estrarre uno o più fattori, ovvero variabili latenti che definiscono marcatamente il costrutto indagato. Se l'analisi porta all'estrazione di un solo fattore principale allora potremmo affermare che le risposte agli item variano allo stesso modo e in maniera coerente al costrutto indagato. Se, invece, osservassimo dei risultati non associati ad uno stesso fattore principale, i relativi item andrebbero scartati dalla scala, in quanto la scala depurata da essi risulterebbe più coerente. Nello specifico del questionario preso in considerazione, si presuppone che lo strumento sia multidimensionale; si tratta dunque di andare a verificare questa caratteristica e identificare in maniera più approfondita quanti siano i fattori sottostanti e da quali elementi siano caratterizzati. Qualora risultassero degli item non direttamente collegati ad alcuno dei componenti estratti, allora essi andrebbero eliminati dalla scala.

Per l'estrazione dei fattori si utilizza il *Metodo delle componenti principali*, per l'individuazione del numero di fattori si utilizza la *Regola degli autovalori maggiori di 1*. Il numero massimo di fattori estraibili è pari al numero di variabili osservate, nel nostro caso 17, ma lo scopo dell'analisi fattoriale consiste nel trovare una o più dimensioni latenti che sintetizzino i dati osservati.

Anche questa analisi è stata condotta su 17 item, non considerando l'item 14 sulla soddisfazione complessiva. Si è utilizzato il metodo di eliminazione pairwise in modo da non precludere l'utilizzo di un questionario nel caso di risposta mancante ad anche uno solo dei 17 item; si è, dunque, considerato il maggior numero di casi possibile.

La Tabella 5 fornisce una rappresentazione sintetica della percentuale di varianza spiegata.

Si osserva che i fattori estratti tramite la *Regola degli autovalori maggiori di 1* sono tre, i quali spiegano complessivamente il **71,294%** della varianza; in particolare il primo fattore da solo riesce a spiegarne il **56,874%**.

Componente	Autovalore	% di varianza	% cumulata
1	9,669	56,874	56,874
2	1,318	7,755	64,628
3	1,133	6,666	71,294

Tabella 5: Analisi fattoriale con 3 fattori

Dalla matrice dei componenti ruotata, ottenuta attraverso la rotazione Varimax e rappresentata in Tabella 6, si osserva che il **primo fattore** si distingue rispetto agli altri per la capacità di spiegare in maniera più adeguata il costrutto preso in considerazione, ovvero la soddisfazione della didattica da parte degli studenti. In particolare a questo componente sono associati tutti gli item che caratterizzano gli indici sintetici Organizzazione Didattica (**item 1, 2, 3 e 8**) e Azione Didattica (**item 6, 7 e 9**). Si possono dunque considerare questi due indici come buoni indicatori della soddisfazione degli studenti. Si osserva, inoltre, che l'item 5 sembra appartenere a questo primo fattore, ma non in maniera rilevante; il coefficiente associato ad esso è infatti pari a 0,422.

Al **secondo fattore** sono legati gli **item 16, 17 e 18**, i quali indagano rispettivamente l'interesse nei confronti dei contenuti dell'insegnamento da parte dello studente, la coerenza dell'insegnamento rispetto agli obiettivi del corso di studi e l'opinione sulle competenze fornite dall'insegnamento in relazione all'ambito lavorativo. Questo fattore dunque sintetizza aspetti non direttamente collegabili agli aspetti organizzativi o all'azione didattica del docente, ma sembra riferirsi alle preferenze del singolo studente e all'opinione sul corso messo in relazione al corso di studi e all'ambito lavorativo.

Il **terzo fattore** è caratterizzato dagli **item 12 e 13**, i quali, quindi, vengono nuovamente considerati non adeguati a valutare il costrutto preso in esame, ovvero la soddisfazione generale dell'attività didattica, ma piuttosto sono legati ad un aspetto di tipo logistico che riguarda la struttura di aule e laboratori. Coerentemente con quanto detto in precedenza, non si sceglie di eliminare questi due quesiti dal questionario, perché nonostante non siano in grado di esprimere la soddisfazione per la didattica, indagano un aspetto di interesse per il progetto rivolto al miglioramento della didattica.

Item	Componente		
	1	2	3
D01 obiettivi	0,757	0,355	
D02 modalità	0,758		
D03 orari	0,720		
D04 numero ore	0,706		
D05 preconoscenze	0,422	0,403	
D06 stimolo	0,688	0,524	
D07 chiarezza	0,753	0,434	
D08 materiale	0,712	0,372	
D09 disponibile	0,785		
D10 ricevimento	0,793	0,360	
D11 esercitazioni	0,687	0,382	0,332
D12 aule			0,914
D13 attrezzature			0,866
D15 carico	0,570	0,349	
D16 interesse	0,384	0,801	
D17 coerente		0,858	
D18 utile		0,834	

Tabella 6: Matrice dei componenti ruotata. Si riportano i coefficienti maggiori di 0,30.

In Tabella 7 sono riportati i risultati dell'analisi fattoriale condotta considerando un numero di fattori fisso e pari a quattro. I quattro componenti estratti spiegano il **75,671%** della varianza, in particolare il quarto fattore contribuisce alla percentuale con il **4,377%**.

Componente	Autovalori	% di varianza	% cumulata
1	9,669	56,874	56,874
2	1,318	7,755	64,628
3	1,133	6,666	71,294
4	0,744	4,377	75,671

Tabella 7: Analisi fattoriale con 4 fattori.

Dalla matrice dei componenti ruotata attraverso la rotazione Varimax (Tabella 8) si osserva che al nuovo fattore estratto sono legati gli item **5** (preconoscenze) e **15** (carico di studio). Ciò

conferma quanto visto in precedenza riguardo al quesito 5, il quale non sembra essere legato al fattore principale. L'analisi, inoltre, conferma le tre dimensioni osservate precedentemente.

Item	Componente			
	1	2	3	4
D01 obiettivi	0,694			
D02 modalità	0,740			
D03 orari	0,775			
D04 numero ore	0,559		0,527	
D05 preconoscenze			0,776	
D06 stimolo	0,590	0,469	0,433	
D07 chiarezza	0,654	0,377	0,440	
D08 materiale	0,603		0,451	
D09 disponibile	0,780			
D10 ricevimento	0,784	0,357		
D11 esercitazioni	0,589		0,432	
D12 aule				0,911
D13 attrezzature				0,856
D15 carico	0,378		0,628	
D16 interesse		0,776		
D17 coerente		0,858		
D18 utile		0,825		

Tabella 8: Matrice dei componenti ruotata con 4 fattori. Si riportano i coefficienti maggiori di 0,35.

3.1.3 Indici di coerenza interna

Il coefficiente *Alpha di Cronbach* è una misura di coerenza interna della scala che si basa sul grado di correlazione media tra item, in particolare indica la proporzione di varianza condivisa dagli item che è attribuibile al costrutto indagato.

Il valore risultante dall'analisi del questionario di soddisfazione è pari a **0,971** e indica un grado di coerenza interna molto alto. In Tabella 9 vengono riportati i valori del coefficiente *Alpha di Cronbach* nel caso in cui sia stato escluso l'item corrispondente. Se l'indice in tabella supera l'*Alpha di Cronbach* calcolato sulla scala nel suo complesso, allora eliminando

quell'item l'affidabilità complessiva della scala aumenterebbe. In questo caso l'*Alpha di Cronbach* relativo alla scala nel suo complesso è pari a 0,971 e l'*Alpha* calcolato sulla scala depurata è uguale a 0,972 in corrispondenza dell'**item 12**; ciò significa che l'item 12 non è sufficientemente coerente con il resto della scala. Anche l'**item 13** sembra non aggiungere molta informazione per quanto riguarda la soddisfazione della didattica. Si ricorda che questi due item presentano, infatti, i valori di correlazione tra item e scala più bassi (vedi Tabella 3). Sembrerebbe, dunque, che i quesiti riguardanti l'adeguatezza delle aule (item 12), dei locali e delle attrezzature per i laboratori (item 13) non siano adatti a valutare adeguatamente la soddisfazione della didattica. Ad essi si aggiunge l'**item 5** in corrispondenza del quale si nota un coefficiente *Alpha di Cronbach* pari a 0,971; anche il quesito relativo all'adeguatezza delle conoscenze preliminari per la comprensione degli argomenti del corso sembrerebbe, dunque, non apportare sufficiente informazione nel valutare la soddisfazione generale della didattica. Tutto ciò conferma quanto già osservato in precedenza tramite l'analisi della correlazione e l'analisi fattoriale.

Item	Alpha di Cronbach se l'item è escluso
D01 obiettivi	0,969
D02 modalità	0,969
D03 orari	0,970
D04 numero ore	0,969
D05 preconoscenze	0,971
D06 stimolo	0,968
D07 chiarezza	0,968
D08 materiale	0,969
D09 disponibile	0,969
D10 ricevimento	0,969
D11 esercitazioni	0,969
D12 aule	0,972
D13 attrezzature	0,971
D15 carico	0,970
D16 interesse	0,969
D17 coerente	0,969
D18 utile	0,970

Tabella 9: Alpha di Cronbach ottenuto escludendo l'item corrispondente.

Per valutare la coerenza interna esistono altri indici che hanno lo scopo di valutare l'affidabilità della scala. Di seguito analizzeremo tre indici diversi: il coefficiente *Split-half R* (Correlazione fra forme), lo *Spearman-Brown Y* e il *Guttman G*. Questi indici si basano sulla divisione della scala a metà. Il principio sottostante è che, se la scala è internamente coerente, allora gruppi distinti di item, ottenuti dividendo a metà la scala di partenza, dovrebbero fornire risultati simili.

Si utilizza nuovamente la scala depurata dall'item 14.

La scala è stata dunque suddivisa in due sottoscale, la prima contenente i 9 item dispari, la seconda contenente i restanti 8 item pari.

Nella parte superiore della Tabella 10 vengono rappresentati i valori dell'Alpha di Cronbach calcolati sulle due sotto-scale. Questi valori risultano più bassi rispetto all'Alpha complessivo, ma comunque molto elevati e tendenzialmente in linea tra loro.

Gli altri tre indici risultano molto elevati e confermano, quindi, una forte coerenza interna.

Alpha di Cronbach	Parte 1	Valore	0,944
		N di item	9
	Parte 2	Valore	0,938
		N di item	8
Numero di item totale			17
Correlazione fra forme			0,971
Coefficiente di Spearman-Brown			0,985
Divisione a metà di Guttman			0,982

Tabella 10: Indici di coerenza interna.

3.1.4 Procedura split-half sample

La procedura *Split-half sample* consiste nella divisione del campione a metà. Questa procedura si basa sul principio secondo il quale per essere affidabile la scala di misurazione deve produrre risultati simili su campioni differenti estratti dalla stessa popolazione o da popolazioni analoghe. Tramite questa procedura si può verificare se gli indici di coerenza interna calcolati nei due sotto-campioni risultano tra loro comparabili.

Il campione complessivo viene dunque suddiviso in due sottogruppi estratti casualmente. In questo caso la numerosità dei dati è pari (163.626 casi), quindi risultano due sotto-campioni con numerosità identica pari a 81.813. Nel primo sotto-campione il numero di casi validi, ovvero che presentano risposta a tutti i 17 quesiti, è pari a 27.433 (33,5% del totale), nel secondo sotto-campione è pari a 27.530 (il 33,6% del totale). In entrambi i sottogruppi l'Alpha di Cronbach risulta pari a **0,971**, equivalentemente a quanto trovato nel campione intero. L'analisi degli indici di coerenza interna in entrambi i sottogruppi porta agli stessi risultati visti nel caso del campione complessivo (vedi Tabella 10).

Si sono dunque ottenuti risultati analoghi su campioni differenti di soggetti estratti dalla stessa popolazione; si può quindi affermare che la scala risulta affidabile.

Attraverso la procedura *split half sample* è, inoltre, possibile eseguire un'analisi a livello di singolo item, verificando che le risposte a ciascun item si comportano in maniera coerente in entrambi i sottogruppi. È stato dunque eseguito un *test sulle differenze* tra i punteggi medi di ciascun item, verificando se questi scostamenti fossero effettivamente significativi o meno. In questo caso si utilizza il *t-test a campioni indipendenti* in quanto si vanno a confrontare valori medi di una variabile (le risposte al singolo item) su due sotto-campioni tra loro indipendenti.

In Tabella 11 si osserva che i valori del *t test* sono ovunque non significativi, i p-value sono infatti tutti maggiori del valore 0,10. Si può concludere che le medie dei punteggi degli item non sono statisticamente differenti tra i due sotto-campioni; in altre parole gli item della scala si comportano in maniera omogenea sui due sottogruppi e ciò denota un forte grado di affidabilità della scala.

Item	Gruppo	Numero questionari	Medie	Statistica t	P-value
D01 obiettivi	1	79.564	7,91	-1,413	0,158
	0	79.609	7,92		
D02 modalità	1	79.098	8,01	1,423	0,155
	0	79.127	7,99		
D03 orari	1	80.219	8,34	0,1	0,920
	0	80.198	8,34		
D04 numero ore	1	73.373	7,70	-1,282	0,200
	0	73.378	7,71		
D05 preconoscenze	1	80.168	7,35	-0,896	0,370
	0	80.210	7,36		
D06 stimolo	1	80.159	7,55	0,128	0,899
	0	80.228	7,55		
D07 chiarezza	1	80.168	7,61	0,13	0,896
	0	80.206	7,61		
D08 materiale	1	79.957	7,49	-0,487	0,626
	0	80.041	7,49		
D09 disponibile	1	79.911	8,11	-0,755	0,450
	0	79.998	8,11		
D10 ricevimento	1	38.862	8,21	0,495	0,620
	0	39.043	8,20		
D11 esercitazioni	1	48.888	7,74	-0,815	0,415
	0	49.183	7,75		
D12 aule	1	80.148	7,52	0,416	0,678
	0	80.179	7,52		
D13 attrezzature	1	50.060	7,53	0,304	0,761
	0	50.075	7,53		
D14 sodd complessiva	1	80.137	7,51	-0,472	0,637
	0	80.135	7,51		
D15 carico	1	80.006	7,32	-1,267	0,205
	0	80.073	7,34		
D16 interesse	1	80.098	7,99	-0,612	0,540
	0	80.102	7,99		
D17 coerente	1	78.737	8,19	-0,753	0,451
	0	78.767	8,20		
D18 utile	1	74.588	7,71	-0,1	0,920
	0	74.563	7,71		

Tabella 11: Test sulla differenza delle medie di due campioni indipendenti.

Il test delle differenze è stato effettuato anche sulla media complessiva degli item: la statistica t è risultata pari a -0,529 e il p-value pari a 0,511; anche in questo caso si accetta l'ipotesi di uguaglianza dei punteggi medi nei due sotto-campioni.

3.2 Verifica della validità del criterio

La verifica della validità del criterio consiste nel valutare se la soddisfazione misurata tramite la scala varia in modo coerente, ovvero nella stessa direzione e con la stessa intensità, con altre variabili chiamate “variabili-criterio”. Per verificare che le misure si muovano nella stessa direzione si utilizza l'*analisi di correlazione*, per quanto riguarda l'intensità si utilizza l'*ANOVA univariata*.

La variabile criterio corrisponde all'item 14, considerato come un indice sintetico del grado di soddisfazione, ottenuto tramite una domanda diretta. Questa variabile viene messa in relazione alle seguenti variabili, costruite tramite trasformazione degli item:

- MEDIA_17 = media aritmetica dei punteggi dei 17 item (escluso item 14)
- ASO = media aritmetica dei punteggi degli item 1, 2, 3 e 8 (aspetti organizzativi)
- AZD = media aritmetica dei punteggi degli item 6, 7 e 9 (azione didattica)

Le correlazioni tra l'item 14 e i tre indicatori risultano tutte e tre statisticamente significative e molto elevate. In particolare la correlazione è pari a :

- 0,875 tra item 14 e MEDIA_17
- 0,797 tra item 14 e ASO
- 0,876 tra item 14 e AZD

Questi valori ci permettono di concludere che la proprietà di validità del criterio è soddisfatta.

Per un'ulteriore conferma è possibile studiare l'intensità delle variabili e, in particolare, verificare se a livelli crescenti di una variabile corrispondano livelli crescenti di un'altra variabile. Per questa analisi si utilizza l'*ANOVA ad una via*.

Le statistiche descrittive riportate in Tabella 12 evidenziano come i valori medi della variabile MEDIA_17 aumentino al crescere del livello della variabile criterio “item 14”. Ciò segnala una correlazione positiva tra le due variabili.

Item 14	Numero questionari	Media
1,0	3.499	3,5796
2,0	2.136	4,7088
3,0	2.599	5,2044
4,0	3.665	5,6767
5,0	7.374	6,1103
6,0	17.186	6,6566
7,0	31.666	7,3660
8,0	39.972	8,0706
9,0	30.113	8,7625
10,0	22.057	9,4751

Tabella 12: Statistiche descrittive della variabile MEDIA_17.

Il valore della statistica F , pari a 59.719,857, e il livello di significatività, circa pari a zero, dimostrano che la varianza della variabile dipendente MEDIA_17 è in gran parte dovuta ad una variabilità del fenomeno tra i gruppi, ovvero è attribuibile alla variabile criterio “item 14”. Le due misure sembrano, dunque, muoversi con la stessa intensità, a conferma di quanto visto in precedenza.

Le stesse osservazioni vengono fatte considerando come variabile dipendente l’indicatore ASO e l’indicatore AZD. I risultati portano alle stesse conclusioni ottenute nel caso della variabile MEDIA_17, ovvero le 10 medie sono statisticamente diverse tra loro e le variabili si muovono con la stessa intensità. Si ha, dunque, un’ulteriore conferma della validità del criterio del questionario adottato.

3.3 Validazione delle due sottoscale: *Aspetti Organizzativi e Azione Didattica*

In questo paragrafo si analizza la validazione dell'affidabilità interna delle due sotto-scale introdotte nel Paragrafo 2.2:

- **aspetti organizzativi**, che raggruppa gli **item 1** (chiarezza sugli obiettivi del corso), **2** (chiarezza sulle modalità d'esame), **3** (rispetto degli orari di lezione) e **8** (adeguatezza del materiale didattico);
- **azione didattica**, che raggruppa gli **item 6** (stimolo/motivazione verso la disciplina), **7** (chiarezza espositiva) e **9** (attenzione verso le esigenze degli studenti).

Il numero di questionari utilizzati per le seguenti analisi è pari a 155.330 per l'indicatore "aspetti organizzativi" e pari a 158.821 per l'indicatore "azione didattica".

3.3.1 Aspetti Organizzativi

L'*Alpha di Cronbach* della scala "Aspetti organizzativi", pari a **0,854**, risulta molto buono, sembra dunque esserci una forte coerenza interna tra gli item che compongono questa sottoscala. Dalla Tabella 13 si nota che tutti gli item sembrano essere coerenti con il costrutto indagato: nessuno degli item, se eliminato, aumenterebbe l'affidabilità interna. Anche i valori delle correlazione tra item e scala complessiva indicano una forte coerenza interna.

Item	Media	Correlazione item-to-total	Alpha di Cronbach se item eliminato
D01 obiettivi	7,91	0,775	0,781
D02 modalità	8,00	0,732	0,798
D03 orari	8,34	0,607	0,849
D08 materiale	7,49	0,677	0,824

Tabella 13: Correlazione item-to-total e Alpha di Cronbach ottenuto se l'item corrispondente viene eliminato (ASO).

Dalla Tabella 14 si osserva che l'analisi fattoriale porta all'estrazione di un fattore, come sperato, che spiega il **69,713%** della varianza.

Componenti	Autovalori	% di varianza	% di varianza cumulata
1	2,789	69,713	69,713
2	0,529	13,236	82,948
3	0,411	10,270	93,218
4	0,271	6,782	100,000

Tabella 14: Analisi fattoriale delle sottoscala "Aspetti Organizzativi".

In Tabella 15 si osserva che tutti i pesi fattoriali, usati per esprimere le variabili originali come combinazione lineare dei fattori, sono molto elevati.

Item	Coefficienti
D01 obiettivi	0,886
D02 modalità	0,861
D03 orari	0,766
D08 materiale	0,821

Tabella 15: Matrice dei componenti (Aspetti Organizzativi).

3.3.2 Azione didattica

Il valore dell'*Alpha di Cronbach* per la sotto-scala "Azione didattica", pari a **0,899**, è ottimo. Dalla Tabella 16 si nota che se escludessimo dalla sottoscala l'**item 9**, relativo alla disponibilità del docente nei confronti degli studenti, l'Alpha di Cronbach aumenterebbe ($0,919 > 0,899$). Questo item risulta quindi non coerente con ciò che la sottoscala vuole indagare, ovvero l'azione didattica. Questi risultati hanno confermato la correttezza della decisione di eliminare questo quesito dal questionario nell'anno accademico successivo (2013/2014). Gli altri due item risultano invece coerenti e presentano valori di correlazione item-to-total elevati.

Item	Media	Correlazione item-to-total	Alpha di Cronbach se item eliminato
D06 stimolo	7,55	0,842	0,819
D07 chiarezza	7,61	0,846	0,815
D09 disponibile	8,11	0,724	0,919

Tabella 16: Correlazione item-to-total e Alpha di Cronbach ottenuto se l'item corrispondente viene eliminato (AZD).

Tramite l'analisi fattoriale viene estratto un fattore che spiega l'**83,233%** della varianza (vedi Tabella 17).

Componenti	Autovalori	% di varianza	% di varianza cumulata
1	2,497	83,233	83,233
2	0,353	11,759	94,992
3	0,150	5,008	100,000

Tabella 17: Analisi fattoriale delle sottoscala "Azione Didattica".

I valori dei pesi fattoriali presenti in Tabella 18 sembrano essere tutti elevati, ma quello corrispondente all'item 9 risulta essere il più basso; ciò sta a indicare che il quesito 9 influisce in maniera minore sul fattore rispetto agli altri due item.

Item	Coefficienti
D06 stimolo	0,933
D07 chiarezza	0,934
D09 disponibile	0,868

Tabella 18: Matrice dei componenti (Azione Didattica).

3.4 Validazione della scala a 4 passi

Come suggerito da ANVUR (Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca) si è deciso di ripetere il processo di validazione della scala, visto nei Paragrafi 3.1 e 3.2 di questo capitolo, dopo aver trasformato le risposte agli item su una scala a 4 passi, anziché a 10 passi come la scala originale. Si è deciso di utilizzare la trasformazione illustrata in Tabella 19 e suggerita nel documento “*Progettazione, implementazione e validazione di un questionario per la valutazione della didattica erogata a studenti universitari*” (2010), realizzato dal Comitato Nazionale per la Valutazione del Sistema Universitario (CNVSU)⁸.

	Molto Insoddisfatto	Insoddisfatto	Soddisfatto	Molto Soddisfatto
Scala originale	1, 2	3, 4, 5	6, 7, 8	9, 10
Scala trasformata	1	2	3	4

Tabella 19: Trasformazione scala a 4 passi.

Di seguito viene presentata la procedura di validazione della scala a 4 passi, mettendola a confronto con la validazione della scala originale a 10 passi. I dati a disposizione sono gli stessi utilizzati nelle precedenti analisi, per cui anche in questo caso l’analisi è stata effettuata sui 163.626 record a disposizione, di cui 54.963 (il 33,6% del totale) non presenta alcuna risposta mancante sui 17 item (escluso l’item 14 sulla soddisfazione generale).

3.4.1 Statistiche descrittive

Dalla Tabella 20 si osserva che le medie dei punteggi degli item vanno da un valore minimo pari a 3,107 in corrispondenza del quesito 15 fino ad un massimo pari a 3,470 relativo al quesito 3. Le medie nella scala a 4 passi seguono, quindi, lo stesso andamento delle medie della scala a 10 passi e in particolare sono tutte vicine al valore massimo pari a 4, confermando un valore medio di soddisfazione molto alto. I valori della deviazione standard variano tra 0,6648 e 0,7705.

⁸ http://www.cnvsu.it/_library/downloadfile.asp?id=11775

Item	Numero questionari	Media	Deviazione standard
D01 obiettivi	159.173	3,312	0,6852
D02 modalità	158.225	3,350	0,7093
D03 orari	160.417	3,470	0,6648
D04 numero ore	146.751	3,245	0,7267
D05 preconnosce	160.378	3,116	0,7193
D06 stimolo	160.387	3,188	0,7705
D07 chiarezza	160.374	3,212	0,7618
D08 materiale	159.998	3,166	0,7444
D09 disponibile	159.909	3,381	0,6899
D10 ricevimento	77.905	3,419	0,6901
D11 esercitazioni	98.071	3,259	0,7341
D12 aule	160.327	3,176	0,7654
D13 attrezzature	100.135	3,184	0,7586
D14 sodd complessiva	160.272	3,170	0,7227
D15 carico	160.079	3,107	0,7577
D16 interesse	160.200	3,342	0,6969
D17 coerente	157.504	3,416	0,6885
D18 utile	149.151	3,246	0,7344

Tabella 20: Statistiche descrittive dei 18 item (scala a 4 passi).

3.4.2 Verifica dell'affidabilità interna

In Tabella 21 si osserva che gli indici di correlazione tra singolo item e totale della scala sono tutti superiori a 0,30. Dal confronto con gli indici emersi dall'analisi sulla scala originale si nota che questi ultimi erano risultati in generale più elevati. Nonostante ciò viene confermata la bassa correlazione tra item e totale della scala in corrispondente dei quesiti numero 12, 13 e 5.

Item	Correlazione <i>item-to-total</i>
D01 obiettivi	0,821
D02 modalità	0,793
D03 orari	0,755
D04 numero ore	0,776
D05 preconoscenze	0,690
D06 stimolo	0,842
D07 chiarezza	0,844
D08 materiale	0,819
D09 disponibile	0,830
D10 ricevimento	0,813
D11 esercitazioni	0,816
D12 aule	0,602
D13 attrezzature	0,650
D15 carico	0,753
D16 interesse	0,794
D17 coerente	0,770
D18 utile	0,748

Tabella 21: Correlazione item-to-total (scala a 4 passi).

Le correlazioni tra gli item, che si possono osservare in Tabella 22, sono tutte inferiori rispetto a quelle ottenute con la scala originale, ma mettono in evidenza gli stessi aspetti. Si nota una forte correlazione tra la soddisfazione complessiva e gli item 6 e 7 che indagano rispettivamente lo stimolo verso la disciplina e la chiarezza nell'esposizione degli argomenti. Gli item 12 e 13 relativi all'adeguatezza delle aule, dei laboratori e delle attrezzature presentano i livelli di correlazione più bassi.

	D01	D02	D03	D04	D05	D06	D07	D08	D09	D10	D11	D12	D13	D15	D16	D17	D18
D02	,660																
D03	,511	,516															
D04	,540	,501	,516														
D05	,428	,364	,310	,421													
D06	,634	,537	,442	,498	,468												
D07	,670	,560	,469	,521	,450	,790											
D08	,591	,533	,444	,511	,438	,632	,682										
D09	,579	,562	,528	,497	,381	,634	,641	,603									
D10	,640	,638	,652	,582	,465	,658	,666	,641	,827								
D11	,590	,549	,499	,624	,474	,625	,624	,614	,626	,715							
D12	,305	,292	,286	,327	,284	,290	,290	,304	,317	,409	,399						
D13	,399	,380	,367	,431	,356	,384	,382	,406	,404	,494	,522	,762					
D15	,481	,448	,382	,547	,432	,519	,519	,519	,489	,540	,552	,320	,412				
D16	,520	,443	,386	,405	,419	,637	,572	,514	,500	,578	,534	,280	,370	,482			
D17	,473	,410	,370	,372	,385	,530	,490	,457	,461	,563	,500	,271	,368	,430	,712		
D18	,464	,389	,325	,380	,366	,517	,482	,454	,421	,514	,518	,266	,372	,422	,621	,712	
D14	,667	,570	,476	,564	,455	,767	,761	,661	,633	,663	,666	,336	,441	,598	,646	,545	,542

Tabella 22: Inter-item correlations (scala a 4 passi).

La Tabella 23 riporta i risultati ottenuti dall'analisi fattoriale effettuata sui 17 item; utilizzando la regola degli autovalori maggiori di 1 sono stati individuati tre componenti i quali spiegano complessivamente il 66,917% della varianza spiegata, di cui il 52,585% è spiegata solo dal primo componente. Il valore della varianza spiegata da questi tre fattori risulta inferiore rispetto al 71,294% riscontrato dall'analisi fattoriale effettuata sulla scala originale.

Componente	Autovalori	% di varianza	% cumulata
1	8,939	52,585	52,585
2	1,301	7,653	60,238
3	1,135	6,679	66,917

Tabella 23: Analisi fattoriale sui 17 item (scala a 4 passi).

Dalla Tabella 24 si osserva che al primo fattore sono legati la maggior parte degli item, in particolare gli item che compongono le due sotto-scale "Azione didattica" e "Organizzazione

didattica”. Come nel caso della scala originale, l’item 5 non sembra appartenere in maniera rilevante a nessuno dei fattori estratti. Il secondo fattore presenta i punteggi più alti ed è caratterizzato, come già visto nel caso della scala originale, da aspetti quali l’interesse verso i contenuti, la coerenza dell’insegnamento rispetto agli obiettivi del corso e le competenze fornite in relazione all’ambito lavorativo. Il terzo fattore è coerente con quanto visto in precedenza, riguarda infatti gli aspetti di tipo logistico.

Item	Componenti		
	1	2	3
D01 obiettivi	0,733		
D02 modalità	0,738		
D03 orari	0,701		
D04 numero ore	0,679		
D05 preconoscenze	0,412	0,371	
D06 stimolo	0,666	0,503	
D07 chiarezza	0,732	0,412	
D08 materiale	0,688	0,352	
D09 disponibile	0,759		
D10 ricevimento	0,774		
D11 esercitazioni	0,665	0,362	
D12 aule			0,905
D13 attrezzature			0,859
D15 carico	0,536	0,350	
D16 interesse	0,365	0,778	
D17 coerente		0,841	
D18 utile		0,817	

Tabella 24: Matrice dei componenti ruotata (scala a 4 passi). Si riportano solo i coefficiente maggiori di 0,35.

La Tabella 25 riporta i risultati dell’analisi fattoriale condotta con numero di fattori fisso e pari a 4. La varianza spiegata da questi 4 fattori messi assieme è del 71,471 % rispetto a quella totale. Il quarto fattore apporta un aumento del 4,6% circa della varianza.

Componente	Autovalori	% di varianza	% cumulata
1	8,939	52,585	52,585
2	1,301	7,653	60,238
3	1,135	6,679	66,917
4	0,774	4,554	71,471

Tabella 25: Analisi fattoriale con 4 componenti (scala a 4 passi).

La matrice dei componenti ruotata visibile in Tabella 26 mette in evidenza il quarto fattore, il quale è formato dagli item 5 e 15, legati rispettivamente alle conoscenze preliminari dello studente e al carico di studio.

Item	Componenti			
	1	2	3	4
D01 obiettivi	0,666			
D02 modalità	0,728			
D03 orari	0,771			
D04 numero ore	0,542		0,486	
D05 preconoscenze			0,778	
D06 stimolo	0,544	0,437	0,472	
D07 chiarezza	0,611		0,474	
D08 materiale	0,560		0,479	
D09 disponibile	0,746			
D10 ricevimento	0,760			
D11 esercitazioni	0,559		0,438	
D12 aule				0,901
D13 attrezzature				0,848
D15 carico			0,613	
D16 interesse		0,754		
D17 coerente		0,847		
D18 utile		0,811		

Tabella 26: Matrice dei componenti ruotata, con 4 componenti (scala a 4 passi). Si riportano solo i coefficienti maggiori di 0,35.

Anche i risultati dell'analisi fattoriale portano, dunque, alle stesse conclusioni viste nel caso della scala originale: il questionario di soddisfazione della didattica da parte degli studenti sembra indagare un costrutto multidimensionale; il fattore principale sembra essere legato agli aspetti organizzativi e all'azione didattica; un fattore è legato ad aspetti logistici, un altro è connesso ai contenuti dell'insegnamento e l'ultimo fattore estratto raggruppa conoscenze preliminari dello studente e carico di studio dell'attività didattica.

L'indice Alpha di Cronbach risulta pari a 0,965 e indica un alto grado di coerenza interna. Questo indice è però inferiore rispetto allo 0,971 riscontrato nel caso di scala a 10 passi.

In Tabella 27 non si osservano valori dell'Alpha di Cronbach superiori a 0,965. A differenza di quanto osservato nella scala a 10 passi, dunque, non risultano item del tutto incoerenti con il resto della scala. Si può però notare che l'item 12, in corrispondenza del quale l'indice Alpha è pari a 0,965, non sembra sufficientemente coerente. Gli item 5 e 13 sembrano essere coerenti, al contrario di quanto visto in precedenza.

Item	Alpha di Cronbach se l'item è escluso
D01 obiettivi	0,962
D02 modalità	0,962
D03 orari	0,963
D04 numero ore	0,962
D05 preconoscenze	0,964
D06 stimolo	0,961
D07 chiarezza	0,961
D08 materiale	0,962
D09 disponibile	0,962
D10 ricevimento	0,962
D11 esercitazioni	0,962
D12 aule	0,965
D13 attrezzature	0,964
D15 carico	0,963
D16 interesse	0,962
D17 coerente	0,962
D18 utile	0,963

Tabella 27: Alpha di Cronbach eliminando ciascun item (scala a 4 passi).

Per la procedura *split-half item* sono stati divisi i 17 item in due scale come nel caso della scala a 10 passi. La prima scala è composta dai nove item codificati con numero dispari e la seconda dagli otto item codificati con numero pari.

Tutti gli indici di coerenza interna presenti in Tabella 28 assumono valori elevati e molto simili nelle due divisioni di scala, confermando l'alto grado di coerenza della scala. Tuttavia questi valori risultano leggermente inferiori rispetto a quelli osservati nel caso della scala originale.

Cronbach's Alpha	Part 1	Value	0,932
		N of Items	9
	Part 2	Value	0,925
		N of Items	8
Total N of Items			17
Correlation Between Forms			0,960
Spearman-Brown Coefficient			0,980
Guttman Split-Half Coefficient			0,977

Tabella 28: : Indici di coerenza interna (scala a 4 passi).

L'Alpha di Cronbach in entrambi i sottogruppi risulta uguale all'indice calcolato sull'intero campione. I valori degli indici di coerenza nei due sottocampioni risultano uguali tra loro e uguali a quelli del campione intero, confermando un forte grado di coerenza interna. Anche i test delle differenze tra i punteggi medi degli item portano a concludere che la scala a 4 passi risulta affidabile.

3.4.3 Verifica della validità del criterio

Le correlazioni tra l'item 14 e i tre indicatori risultano tutte e tre statisticamente significative e molto simili a quelle calcolate sulla scala a 10 passi. In particolare i valori risultano pari a :

- 0,823 tra item 14 e MEDIA_17
- 0,736 tra item 14 e ASO
- 0,811 tra item 14 e AZD

Anche le tre Analisi della Varianza effettuate considerando l'item 14 come variabile indipendente e le variabili MEDIA_17, ASO e AZ come variabili dipendenti portano a concludere che il questionario a 4 passi è valido.

In conclusione le analisi svolte considerando la scala a 4 passi non si discostano di molto rispetto a quelle effettuate sulla scala originale a 10 passi. Si può, dunque affermare che il questionario di valutazione della didattica da parte degli studenti:

- è affidabile;
- gode della proprietà di validità del criterio;
- non è unidimensionale, ma indaga diversi aspetti. In particolare sono stati individuati, tramite l'analisi fattoriale, quattro dimensioni: una legata alla soddisfazione nei confronti dell'attività didattica, una legata alla soddisfazione nei confronti dei contenuti dell'insegnamento, una riguardante aspetti logistici e l'ultima connessa alle preconoscenze dello studente e al carico di studio.

CAPITOLO 4

I MODELLI A CLASSI LATENTI

Il protocollo tradizionale, di cui si è parlato nel primo capitolo, considera le variabili come quantitative (o metriche) nonostante le scale di misura siano raramente composte da variabili di questo tipo. Gli item che compongono i questionari, infatti, spesso generano variabili di tipo ordinale o nominale. Per questo motivo si è deciso di utilizzare anche l'approccio con modelli a classi latenti per valutare la validità e la dimensionalità della scala di misura. Essi, infatti, oltre a considerare esplicitamente che le variabili generate dagli item sono di tipo categoriale (ordinale o nominale), tengono conto del fatto che il costrutto da misurare è indirettamente osservabile (latente). Inoltre, i modelli a classi latenti non richiedono l'assunzione, spesso violata, di normalità multivariata (McCutcheon, 1987).

In questo capitolo vengono presentati i modelli a classi latenti che verranno poi applicati ai dati relativi al questionario sulla soddisfazione degli studenti.

I modelli a classi latenti furono inizialmente introdotti da Lazarsfeld ed Henry (1968) per misurare variabili latenti attitudinali a partire da item dicotomici. Come già accennato sopra, questi modelli, al contrario dell'analisi fattoriale che considera le variabili come quantitative, tengono conto del loro aspetto qualitativo. In seguito, la metodologia dei modelli a classi latenti fu formalizzata da Goodman (1974a, 1974b), il quale ne estese l'applicazione anche a variabili nominali. Goodman, inoltre, sviluppò l'algoritmo di stima di massima verosimiglianza, sul quale tuttora si basano molti software utilizzati per l'analisi a classi latenti. In anni recenti sono state introdotte delle estensioni per includere variabili osservabili di diverso tipo presenti nella stessa scala di misura (nominale, ordinale, continuo e conteggio) e covariate.

Nei modelli a classi latenti esistono due tipi di variabili: variabili direttamente osservabili, chiamate *indicatori* o *variabili manifeste* o *variabili osservate*, e variabili non direttamente osservabili, chiamate *variabili latenti*; le categorie delle variabili latenti sono chiamate *classi latenti*.

Lo studio delle variabili latenti si basa su una premessa fondamentale: la covariazione osservata tra gli indicatori è dovuta alle relazioni che essi hanno con la variabile latente, cioè è

la variabile latente che spiega le relazioni esistenti tra le variabili osservate. Si studiano, perciò, le relazioni tra le variabili osservabili con lo scopo di analizzare e descrivere le caratteristiche della variabile latente, la quale costituisce il vero oggetto di interesse.

L'analisi a classi latenti può essere usata come tecnica esplorativa per ridurre un insieme di variabili osservate in una o più variabili latenti, costituite da un insieme di classi, oppure come metodo confermativo per verificare ipotesi sulle asserzioni fatte a priori riguardo le relazioni tra le variabili osservate.

In questo capitolo verranno introdotti e descritti tre importanti casi di modelli a classi latenti:

- il modello a classi latenti tradizionale (*latent class cluster model*);
- il modello fattoriale a classi latenti (*latent class factor model*);
- il modello di regressione a classi latenti (*latent class regression model*).

4.1 I modelli a classi latenti tradizionali

L'analisi a classi latenti tradizionale (Goodman, 1974a) assume che ogni osservazione, ovvero il punteggio assegnato dai rispondenti agli item di un questionario, appartenga ad una e una sola classe latente t (con $t = 1, 2, \dots, T$). La relazione tra la variabile latente e i suoi indicatori non è deterministica, ma probabilistica. Inoltre, si assume che esista indipendenza locale (*local independence*) tra le variabili manifeste (indicatori), ovvero condizionatamente all'appartenenza ad una certa classe latente, le variabili manifeste sono mutuamente indipendenti l'una dall'altra: tutto ciò che esse hanno in comune è il fatto di essere indicatori della stessa variabile latente. Le variabili manifeste sono associate tra loro, ma questa associazione sparisce quando la variabile latente è posta costante.

Il modello tradizionale può essere espresso usando come parametri le probabilità marginali di appartenere ad ogni classe latente e le probabilità di risposta condizionate.

Si suppone, per esempio, un modello con quattro variabili nominali manifeste, chiamate A , B , C e D , e una variabile latente, chiamata X , esso è definito come:

$$\pi_{ijkl} = \sum_{t=1}^T \pi_{ijklt}^{ABCDX} \quad (2)$$

con

$$\pi_{ijklt}^{ABCDX} = \pi_t^X \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{lt}^{D|X} \quad (3)$$

dove:

- π_{ijkl} è la probabilità di un individuo di essere al livello (i, j, k, l) rispetto alla variabile congiunta (A, B, C, D) ;
- π_{ijklt}^{ABCDX} indica la probabilità di individuo di essere al livello (i, j, k, l, t) rispetto alla variabile congiunta (A, B, C, D, X) ;
- π_t^X è la probabilità di appartenere alla classe latente $t = 1, 2, \dots, T$ della variabile X ;
- $\pi_{it}^{A|X}$ è la probabilità condizionata di ottenere la i -esima risposta, con $i = 1, 2, \dots, I$, all'item A da parte degli individui appartenenti alla classe latente t ;
- $\pi_{jt}^{B|X}$, $\pi_{kt}^{C|X}$ e $\pi_{lt}^{D|X}$ sono le probabilità condizionate per gli item B , C e D , rispettivamente; con $j = 1, 2, \dots, J$, $k = 1, 2, \dots, K$ e $l = 1, 2, \dots, L$.

Data la variabile latente X , le risposte agli indicatori A , B , C e D osservate sono mutuamente indipendenti come implica l'assunzione di *local independence*. In particolare, l'equazione (2) afferma che gli individui possono essere classificati in T classi latenti esaustive e mutuamente esclusive, e l'equazione (3) indica che all'interno delle classi latenti le variabili osservate sono mutuamente indipendenti.

Il modello descritto in (2) e (3) può essere rappresentato graficamente tramite un diagramma (*path diagram*) nel quale le variabili manifeste non sono connesse direttamente l'una all'altra, ma indirettamente attraverso la fonte comune, ovvero la variabile latente X (Figura 1).

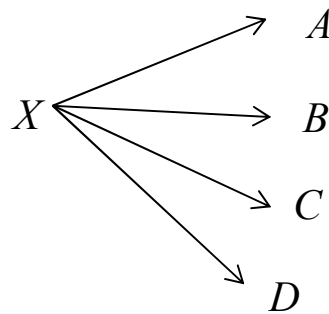


Figura 1: Path diagram di un modello con una variabile latente e quattro indicatori.

Uno degli scopi dei modelli a classi latenti è quello di determinare il minor numero di classi latenti T , in grado di spiegare le associazioni tra le variabili osservabili. L'analisi solitamente inizia con la stima del modello base (H_0) con $T = 1$:

$$\pi_{ijkl}^{ABCD} = \pi_i^A \pi_j^B \pi_k^C \pi_l^D \quad (4)$$

Se questo modello non si adatta adeguatamente i dati, allora si deve stimare un modello con $T=2$ e così via finché non si trova il modello più semplice con il miglior adattamento ai dati.

Haberman (1979) dimostra che il modello a classi latenti descritto in (2) e (3) è equivalente ad un modello log-lineare con variabili latenti e si esprime nel modo seguente:

$$\ln F_{ijklt}^{ABCDX} = \lambda + \lambda_t^X + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{it}^{AX} + \lambda_{jt}^{BX} + \lambda_{kt}^{CX} + \lambda_{lt}^{DX} \quad (5)$$

dove:

- F_{ijklt}^{ABCDX} è la frequenza assoluta della generica cella nella tabella di contingenza a cinque vie;
- $\lambda_t^X, \lambda_i^A, \lambda_j^B, \lambda_k^C$ e λ_l^D sono gli effetti di primo ordine;
- $\lambda_{it}^{AX}, \lambda_{jt}^{BX}, \lambda_{kt}^{CX}$ e λ_{lt}^{DX} sono gli effetti di secondo ordine.

L'assunzione di indipendenza locale è mantenuta omettendo tutti i termini di interazione che si riferiscono alle associazioni tra indicatori.

La relazione tra i parametri nelle due diverse parametrizzazioni, nel caso della variabile osservata A , è:

$$\pi_{it}^{A|X} = \frac{\exp(\lambda_i^A + \lambda_{it}^{AX})}{\sum_{r=1}^I \exp(\lambda_r^A + \lambda_{rt}^{AX})}$$

Quando gli indicatori sono ordinali, l'effetto di secondo ordine in equazione (5) diventa $\lambda_{it}^{AX} = \lambda_t^X i$, dove i è il punteggio assegnato all'item A .

4.1.1 Stima dei parametri del modello

La procedura di stima di massima verosimiglianza dei parametri del modello a classi latenti fu inizialmente delineata da Goodman (1974a, 1974b, 1979). Questa procedura fornisce stime delle probabilità condizionate e marginali consistenti, asintoticamente efficienti e che non possono trovarsi al di fuori dell'intervallo consentito (0,1).

Si inizia modificando l'equazione (3) ponendo l'accento circonflesso ai parametri:

$$\hat{\pi}_{ijklt}^{ABCDX} = \hat{\pi}_t^X \hat{\pi}_{it}^{A|X} \hat{\pi}_{jt}^{B|X} \hat{\pi}_{kt}^{C|X} \hat{\pi}_{lt}^{D|X} \quad (6)$$

In questo modo i parametri del modello sono stati sostituiti con le rispettive stime di massima verosimiglianza. Se si somma l'equazione (6) su tutte le T classi della variabile latente X , si ottiene la stima di massima verosimiglianza della probabilità congiunta associata ad ognuno dei (I, J, K, L) livelli delle variabili osservate:

$$\hat{\pi}_{ijkl} = \sum_{t=1}^T \hat{\pi}_{ijklt}^{ABCDX} \quad (7)$$

Se si divide l'equazione (6) dall'equazione (7), si ottiene la stima di massima verosimiglianza della probabilità condizionate che un'osservazione al livello (i, j, k, l) delle variabili osservate sia al livello t della variabile latente X :

$$\hat{\pi}_{tijkl}^{X|ABCD} = \frac{\hat{\pi}_{ijklt}^{ABCDX}}{\hat{\pi}_{ijkl}} \quad (8)$$

Questa equazione fornisce la probabilità che le osservazioni nella cella (i, j, k, l) della tabella di contingenza a 4 vie siano al livello t della variabile latente X .

Goodman (1974a) nota che se si è in presenza di un insieme di proporzioni p_{ijkl} di individui che sono al livello (i, j, k, l) di una tabella di contingenza delle variabili osservate $(A, B, C$ e $D)$, è possibile utilizzare metodi standard per provare che le stime di massima verosimiglianza dei parametri delle equazioni (6) e (7) soddisfano il seguente insieme di equazioni:

$$\hat{\pi}_t^X = \sum_{ijkl} p_{ijkl} \hat{\pi}_{tijkl}^{X|ABCD} \quad (9)$$

$$\hat{\pi}_{it}^{A|X} = \frac{\sum_{jkl} p_{ijkl} \hat{\pi}_{tijkl}^{X|ABCD}}{\hat{\pi}_t^X} \quad (10)$$

$$\hat{\pi}_{jt}^{B|X} = \frac{\sum_{ikl} p_{ijkl} \hat{\pi}_{tijkl}^{X|ABCD}}{\hat{\pi}_t^X} \quad (11)$$

$$\hat{\pi}_{kt}^{C|X} = \frac{\sum_{ijl} p_{ijkl} \hat{\pi}_{tijkl}^{X|ABCD}}{\hat{\pi}_t^X} \quad (12)$$

$$\hat{\pi}_{lt}^{D|X} = \frac{\sum_{ijk} p_{ijkl} \hat{\pi}_{tijkl}^{X|ABCD}}{\hat{\pi}_t^X} \quad (13)$$

Questo insieme di equazioni viene utilizzato per trovare le stime di massima verosimiglianza delle probabilità condizionate e marginali, attraverso il processo iterativo chiamato EM (Expectation – Maximization).

Siano $\bar{\pi}_t^X$, $\bar{\pi}_{it}^{A|X}$, $\bar{\pi}_{jt}^{B|X}$, $\bar{\pi}_{kt}^{C|X}$ e $\bar{\pi}_{lt}^{D|X}$ i valori iniziali dei parametri, allora si ottiene un valore iniziale $\bar{\pi}_{ijkl}^{ABCDX}$ per $\hat{\pi}_{ijkl}^{ABCDX}$ tramite l'equazione:

$$\bar{\pi}_{ijkl}^{ABCDX} = \bar{\pi}_t^X \bar{\pi}_{it}^{A|X} \bar{\pi}_{jt}^{B|X} \bar{\pi}_{kt}^{C|X} \bar{\pi}_{lt}^{D|X} \quad (14)$$

Questo valore è usato per ottenere i due valori iniziali $\bar{\pi}_{ijkl}$ e $\bar{\pi}_{tijkl}^{X|ABCD}$ tramite:

$$\bar{\pi}_{ijkl} = \sum_{t=1}^T \bar{\pi}_{ijkl}^{ABCDX} \quad (15)$$

$$\bar{\pi}_{ijkl}^{ABCDX} = \frac{\bar{\pi}_{ijkl}^{ABCDX}}{\bar{\pi}_{ijkl}} \quad (16)$$

Usando le proporzioni osservate p_{ijkl} si ottiene un nuovo valore per la stima della probabilità marginale di appartenere alla classe latente t della variabile latente X , $\bar{\pi}_t^X$, attraverso l'equazione:

$$\bar{\pi}_t^X = \sum_{ijkl} p_{ijkl} \bar{\pi}_{tijkl}^{X|ABCD} \quad (17)$$

Questo valore può a sua volta essere utilizzato per ottenere i nuovi valori di stima delle probabilità condizionate $\bar{\pi}_{it}^{A|X}$, $\bar{\pi}_{jt}^{B|X}$, $\bar{\pi}_{kt}^{C|X}$ e $\bar{\pi}_{lt}^{D|X}$, attraverso le equazioni:

$$\bar{\pi}_{it}^{A|X} = \frac{\sum_{jkl} p_{ijkl} \bar{\pi}_{tijkl}^{X|ABCD}}{\bar{\pi}_t^X} \quad (18)$$

$$\bar{\pi}_{jt}^{B|X} = \frac{\sum_{ikl} p_{ijkl} \bar{\pi}_{tijkl}^{X|ABCD}}{\bar{\pi}_t^X} \quad (19)$$

$$\bar{\pi}_{kt}^{C|X} = \frac{\sum_{ijl} p_{ijkl} \bar{\pi}_{tijkl}^{X|ABCD}}{\bar{\pi}_t^X} \quad (20)$$

$$\bar{\pi}_{lt}^{D|X} = \frac{\sum_{ijk} p_{ijkl} \bar{\pi}_{tijkl}^{X|ABCD}}{\bar{\pi}_t^X} \quad (21)$$

Riassumendo, dunque, si inizia con i valori iniziali delle stime delle probabilità condizionate e delle probabilità marginali di appartenere ad una classe latente t . La stima che si ottiene nel lato sinistro dell'equazione (14) viene poi usata nelle equazioni (15) e (16); a sua volta, le stime ottenute da queste due equazioni sono utilizzate per ottenere le nuove stime di massima verosimiglianza tramite le equazioni (17), (18), (19), (20) e (21). Infine, con queste ultime stime ricomincia il processo iterativo.

Esistono due metodi per arrestare questa procedura. Il primo metodo prevede che si possa decidere a priori il numero di iterazioni da completare; lo svantaggio di questo approccio consiste nel fatto che le stime possano cambiare in maniera significativa da un'iterazione all'altra. Il secondo metodo consiste nel confrontare le probabilità ottenute dalle equazioni (17), (18), (19), (20) e (21), con quelle dell'equazione (14), al termine di ogni iterazione; se la differenza tra le stime è inferiore ad un predeterminato valore (*tolerance*), allora il processo si arresta e l'equazione (14) può essere usata per testare l'adattamento del modello ai dati.

Devono essere fatte tre osservazioni riguardo la stima delle probabilità condizionate e marginali:

- 1) Dalle equazioni di verosimiglianza potrebbe risultare più di una soluzione, ovvero potrebbe esistere più di un insieme di probabilità (condizionate e marginali) per ciascun numero di classi T fissato. In questo caso si dovrebbe provare più di un insieme di valori iniziali nell'equazione (14). Nella pratica questo non risulta essere un problema perché, in genere, ogni insieme di valori iniziali porta agli stessi valori di stima finali.
- 2) La seconda osservazione è che il numero di parametri stimabili è limitato dai gradi di libertà disponibili nella tabella di contingenza delle variabili osservate (g.d.l. pari a $(IJKL) - 1$). Nel modello specificato nell'equazione (14), che supponiamo essere senza restrizioni (*unrestricted*) e con tutti i parametri identificati, il numero di parametri stimati è pari a $(T - 1) + T(I - 1) + T(J - 1) + T(K - 1) + T(L - 1) = (I + J + K + L - 3)T - 1$. Solo se il numero di parametri da stimare è inferiore ai gradi di libertà della tabella di contingenza si può stimare il modello specificato nell'equazione (14).
- 3) Non sempre è possibile ottenere stime uniche dei parametri. Una condizione necessaria e sufficiente per determinare l'identificabilità locale (*local identifiability*) di un modello a classi latenti è fornita da Goodman (1974a). Essa consiste nel determinare il rango della matrice di dimensione $[(IJKL) - 1] \times (I + J + K + L -$

3) $T - 1$] delle derivate parziali delle probabilità π_{ijkl} non ridondanti rispetto ai parametri distinti del modello:

$$\begin{pmatrix} \frac{\partial \pi_{1111}}{\partial \pi_1^X} & \cdots & \frac{\partial \pi_{1111}}{\partial \pi_{(L-1),T}^{D|X}} \\ \vdots & \ddots & \vdots \\ \frac{\partial \pi_{IJKL-1}}{\partial \pi_1^X} & \cdots & \frac{\partial \pi_{1111}}{\partial \pi_{(L-1),T}^{D|X}} \end{pmatrix}$$

dove $\frac{\partial \pi_{ijkl}}{\partial \pi_t^X} = \pi_{it}^{A|X} \pi_{jt}^{B|X} \pi_{kt}^{C|X} \pi_{lt}^{D|X} - \pi_{iT}^{A|X} \pi_{jT}^{B|X} \pi_{kT}^{C|X} \pi_{lT}^{D|X}$, ovvero l'elemento di ogni cella è la derivata parziale dell'elemento di riga rispetto a quello della colonna corrispondente.

Perché la condizione di identificabilità locale sia soddisfatta la matrice deve avere rango pieno rispetto alle proprie colonne, ovvero il rango deve essere uguale a $(I + J + K + L - 3)T - 1$. In altre parole non devono esserci colonne linearmente dipendenti. I modelli non identificabili possono essere identificati ponendo alcune restrizioni sui parametri; ciò comporta una diminuzione del numero di parametri distinti di un ammontare pari al numero di restrizioni.

4.1.2 Valutazione dell'adattamento delle stime del modello

Esistono diversi approcci complementari per la valutazione della bontà del modello. Di seguito ne verranno descritti tre.

- 1) Nell'approccio più diffuso si utilizza la statistica rapporto di verosimiglianza L^2 per valutare in che misura le stime di massima verosimiglianza delle frequenze attese di ogni cella \hat{F}_{ijkl}^{ABCD} differiscono dalle corrispondenti frequenze osservate f_{ijkl}^{ABCD} :

$$L^2 = 2 \sum_{ijkl} f_{ijkl}^{ABCD} \ln(\hat{F}_{ijkl}^{ABCD} / f_{ijkl}^{ABCD})$$

Secondo questo criterio, un modello si adatta bene ai dati se il valore della statistica L^2 è sufficientemente basso da essere attribuibile al caso, cioè se la differenza tra

frequenze attese e osservate non è significativa entro i limiti dell'errore statistico, generalmente posto ad un livello pari a 5%.

Le frequenze attese \hat{F}_{ijkl}^{ABCD} sono calcolate secondo una procedura a due passi. Per prima cosa si ottengono le stime di massima verosimiglianza dei parametri del modello e si sostituiscono nel lato destro dell'equazione (3) per ottenere le stime di massima verosimiglianza $\hat{\pi}_{ijkl}^{ABCDX}$. Queste stime vengono poi sommate rispetto alle classi latenti, in modo da ottenere le probabilità stimate per ogni cella della tabella di contingenza, e moltiplicate per la numerosità campionaria N. Si ottengono, quindi, le stime di massima verosimiglianza per le frequenze attese:

$$\hat{F}_{ijkl}^{ABCD} = N \sum_{t=1}^T \hat{\pi}_{ijkl}^{ABCDX}$$

Nel caso in cui le frequenze attese siano uguali alle frequenze osservate ($\hat{F}_{ijkl}^{ABCD} = f_{ijkl}^{ABCD}$) il modello risulterebbe perfetto e L^2 risulterebbe pari a 0. Quando $L^2 > 0$, il valore L^2 misura la mancanza di adattamento del modello, quantificando l'ammontare di associazione che rimane non spiegata dal modello. Quando N è sufficientemente grande e sussistono alcune condizioni di regolarità, L^2 segue una distribuzione χ^2 e, come regola generale, il numero di gradi di libertà (df) è pari al numero di celle nella tabella di contingenza, meno il numero di parametri distinti M, meno uno. Per esempio, nel caso di quattro variabili categoriali il numero di celle nella tabella di contingenza è pari a $IJKL$ e il numero di parametri è $M = T - 1 + T[(I - 1) + (J - 1) + (K - 1) + (L - 1)]$.

Nel caso di dati sparsi⁹ non si dovrebbe usare la distribuzione χ^2 per calcolare il p-value perché la statistica L^2 non risulterebbe ben approssimata, ma si può ricorrere all'uso di un approccio bootstrap (Langeheine, Pannekoek e Van de Pol, 1996).

⁹ Si hanno dati sparsi quando il numero di variabili osservate o il numero di modalità delle variabili è elevato. In questo caso il numero totale di celle nella tabella di contingenza risulta molto più grande rispetto la numerosità del campione, ciò comporta un numero elevato di celle vuote o con frequenze molto basse. Si hanno dati sparsi anche quando i modelli a classi latenti sono estesi alle variabili continue.

- 2) Un altro metodo per valutare l'adattamento del modello nel caso di dati sparsi consiste nell'utilizzo di criteri informativi che tengano conto sia della bontà del modello sia del numero di parametri (parsimonia). Queste misure, come il criterio di Akaike (*AIC*) o il criterio Bayesiano (*BIC*), sono utili per effettuare confronti tra modelli. Nell'analisi a classi latenti si usa generalmente la statistica *BIC*, che è definita come:

$$BIC_L^2 = L^2 - \ln(N) df$$

dove *df* sono i gradi di libertà della statistica L^2 .

Una definizione della statistica *BIC* è basata sulla log-verosimiglianza (*LL*) e il numero di parametri (*M*) e corrisponde a:

$$BIC_{LL} = -2LL + \ln(N) M$$

Un modello che presenta un valore della statistica *BIC* inferiore è preferibile rispetto ad un modello a cui corrisponde un *BIC* maggiore.

- 3) Se il modello di base H_0 , ovvero il modello con $T=1$, fornisce un buon adattamento ai dati, allora non è necessaria nessuna analisi a classi latenti, poiché non c'è nessuna associazione tra le variabili che deve essere spiegata. Tuttavia, nella maggioranza dei casi, il modello base non risulta adatto; in questo caso $L^2(H_0)$ può fornire una misura base dell'associazione nei dati. Il terzo metodo per valutare l'adattamento ai dati consiste nel confrontare i valori della statistica L^2 associata ai diversi modelli a classi latenti con $T > 1$, con il valore $L^2(H_0)$. In questo modo è possibile calcolare la percentuale di riduzione di L^2 , la quale rappresenta l'associazione totale spiegata dal modello. Questo approccio meno formale può integrare metodi statisticamente più precisi come le statistiche L^2 e *BIC*.

4.1.3 La significatività degli effetti

Il passo successivo dell'analisi a classi latenti tradizionale consiste nell'eliminare dal modello tutte quelle variabili che non sembrano variare in maniera significativa tra classi. Per esempio, se si volesse verificare se eliminare la variabile *A* da un modello con *T* classi latenti, è

necessario testare l'ipotesi nulla che la distribuzione delle I categorie della variabile A sia la stessa entro ciascuna classe latente t :

$$\pi_{i1}^{A|X} = \pi_{i2}^{A|X} = \dots = \pi_{iT}^{A|X} \quad \text{per } i = 1, 2, \dots, I$$

Per implementare questo test si utilizza la relazione tra le probabilità di risposta condizionata e i parametri log-lineari:

$$\pi_{it}^{A|X} = \frac{\exp(\lambda_i^A + \lambda_{it}^{AX})}{\sum_{r=1}^I \exp(\lambda_r^A + \lambda_{rt}^{AX})}$$

L'ipotesi nulla espressa in termini dei parametri log-lineari risulta:

$$\lambda_{i1}^{AX} = \lambda_{i2}^{AX} = \dots = \lambda_{iT}^{AX} = \mathbf{0} \quad \text{per } i = 1, 2, \dots, I$$

Un modo per testare questa ipotesi consiste nell'uso della differenza delle statistiche L^2 . La nuova statistica ΔL^2 è calcolata come differenza tra le statistiche L^2 ottenute sotto il modello con e senza restrizioni.

4.1.4 Classificazione

Nell'ultimo passo dell'analisi a classi latenti si utilizzano i risultati del modello per classificare i soggetti (o casi) nelle classi latenti appropriate. Usando il teorema di Bayes si possono ottenere, per ogni schema di risposte (i, j, k, l) , le stime delle probabilità a posteriori di appartenere ad una classe latente, ovvero:

$$\hat{\pi}_{tijk}^{X|ABCD} = \frac{\hat{\pi}_{ijkl}^{ABCDX}}{\sum_{t=1}^T \hat{\pi}_{ijkl}^{ABCDX}}, \quad t = 1, 2, \dots, T \quad (22)$$

dove il numeratore e il denominatore sono ottenuti sostituendo ai parametri del modello (3) le corrispondenti stime.

Magidson e Vermunt (2001) e Vermunt e Magidson (2002) si riferiscono ai modelli a classi latenti tradizionali con il nome modelli a classi latenti di tipo *cluster* perché lo scopo di

classificare i soggetti in T gruppi omogenei è uguale a quello della *cluster analysis*. La differenza tra le due tecniche sta nel fatto che l'analisi a classi latenti definisce la distanza tra classi in termini di probabilità, mentre la *cluster analysis* utilizza una misura di distanza ad hoc. Come indicato nell'equazione (3), i casi che appartengono alla stessa classe latente sono tra loro simili perché le loro risposte sono generate dalla stessa distribuzione di probabilità. I casi sono, dunque, assegnati alla classe che presenta probabilità a posteriori (equazione 22) più elevata.

4.2 I modelli a classi latenti non tradizionali

Può succedere che un modello a classi latenti con T classi sia rifiutato per mancanza di adeguatezza, ciò significa che l'ipotesi di indipendenza locale non è valida. In questo caso la strategia tradizionale prevede la stima di un modello con $T+1$ classi latenti. Si possono, però, adottare strategie alternative che possono portare alla stima di modelli più parsimoniosi rispetto ai modelli tradizionali e più congruenti alle ipotesi iniziali. Queste strategie alternative consistono nell':

- Aggiungere uno o più effetti diretti, che tengano conto dell'associazione residua tra le variabili osservate, responsabili della dipendenza locale. Questo metodo risulta utile quando alcuni fattori esterni, non collegati alla variabile latente, creano un'associazione tra due variabili non pertinente.
- Eliminare uno o più item, responsabili della dipendenza locale. Questa strategia è particolarmente utile quando si è in presenza di molte variabili ridondanti.
- Aumentare il numero di variabili latenti. Questa alternativa è particolarmente utile quando un gruppo di variabili è responsabile della dipendenza locale. Magidson e Vermunt (2001) mostrano che il modello a classi latenti fattoriale (*LC factor model*), ottenuto aumentando il numero di variabili latenti piuttosto che il numero di classi latenti, spesso si adatta meglio ai dati rispetto al modello a classi latenti di tipo cluster (*LC cluster model*) con lo stesso numero di parametri. Inoltre, il modello a classi latenti fattoriale è identificato in alcune situazioni in cui il modello tradizionale non lo è.

La statistica diagnostica BVR (*bivariate residual*) aiuta a definire le relazioni bivariate che il modello non riesce a spiegare adeguatamente e a determinare quale delle tre strategie implementare. Anche in situazioni nelle quali la statistica L^2 indica un buon adattamento complessivo ai dati, l'adattamento rispetto ad una o più tabelle a doppia entrata potrebbe non esserlo. Il BVR misura l'associazione di due variabili osservate spiegata dal modello. Esso corrisponde ad una statistica χ^2 (divisa per i gradi di libertà) in cui le frequenze osservate nella tabella a doppia entrata vengono confrontate con le frequenze attese stimate tramite il modello a classi latenti. Un valore della statistica BVR molto maggiore ad 1 indica che il modello non è in grado di spiegare adeguatamente l'associazione tra le due variabili prese in considerazione.

4.2.1 Modelli a classi latenti fattoriali

Magidson e Vermunt (2001) propongono il modello a classi latenti fattoriale come un'alternativa al modello a classi latenti di tipo esplorativo tradizionale. I modelli a classi latenti di tipo fattoriale permettono di includere più di una variabile nel modello. Per esempio un modello a classi latenti tradizionale con quattro o più classi può essere interpretato in termini di due o più variabili latenti. Una variabile latente X con $T=4$ classi può essere espressa in termini di due variabili latenti dicotomiche $V=\{1,2\}$ e $W=\{1,2\}$, usando la seguente corrispondenza:

	$W=1$	$W=2$
$V=1$	$X=1$	$X=2$
$V=2$	$X=3$	$X=4$

$X=1$ corrisponde a $V=1$ e $W=1$, $X=2$ a $V=1$ e $W=2$, $X=3$ a $V=2$ e $W=1$ e $X=4$ a $V=2$ e $W=2$.

Formalmente, per quattro variabili osservate, il modello a classi latenti con quattro classi può essere riparametrizzato come un modello a classi latenti fattoriale con due variabili latenti dicotomiche:

$$\pi_{ijklrs}^{ABCDVW} = \pi_{rs}^{VW} \pi_{ijklrs}^{ABCD|VW} = \pi_{rs}^{VW} \pi_{irs}^{A|VW} \pi_{jrs}^{B|VW} \pi_{krs}^{C|VW} \pi_{lrs}^{D|VW}$$

dove :

- π_{rs}^{VW} è la probabilità che V e W assumano rispettivamente i valori $r=1,2,\dots,R$ e $s=1,2,\dots,S$;
- $\pi_{irs}^{A|VW}$ la probabilità condizionata che all'item A si risponda i , dato che $V=r$ e $W=s$.

In forma log-lineare:

$$\ln F_{ijklrs}^{ABCDVW} = \lambda + \lambda_r^V + \lambda_s^W + \lambda_{rs}^{VW} + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{ir}^{AV} + \lambda_{jr}^{BV} + \lambda_{kr}^{CV} + \lambda_{lr}^{DV} + \lambda_{is}^{AW} + \lambda_{js}^{BW} + \lambda_{ks}^{CW} + \lambda_{ls}^{DW} + \lambda_{irs}^{AVW} + \lambda_{jrs}^{BVW} + \lambda_{krs}^{CVW} + \lambda_{lrs}^{DVW} \quad (23)$$

La corrispondenza tra le equazioni (5) e (23) è data da:

$$\lambda_{2(r-1)+s}^X = \lambda_r^V + \lambda_s^W + \lambda_{rs}^{VW}$$

$$\lambda_{i,2(r-1)+s}^{AX} = \lambda_{ir}^{AV} + \lambda_{is}^{AW} + \lambda_{irs}^{AVW} \quad \text{ecc.}$$

Magidson e Vermunt (2001) considerano vari modelli a classi latenti fattoriali con restrizioni. Con il termine *basic LC factor models* si riferiscono a modelli a classi latenti che contengono due o più variabili latenti dicotomiche mutuamente indipendenti tra loro e che escludono interazioni di ordine superiore dalle probabilità di risposta condizionate. Un modello fattoriale a classi latenti *basic* si ottiene imponendo appropriate restrizioni al modello a classi latenti fattoriale senza restrizioni, espresso in (23). Nel caso di un modello con due fattori latenti discreti l'espressione è la seguente:

$$\ln F_{ijklrs}^{ABCDVW} = \lambda + \lambda_r^V + \lambda_s^W + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_l^D + \lambda_{ir}^{AV} + \lambda_{jr}^{BV} + \lambda_{kr}^{CV} + \lambda_{lr}^{DV} + \lambda_{is}^{AW} + \lambda_{js}^{BW} + \lambda_{ks}^{CW} + \lambda_{ls}^{DW}$$

nella quale sono stati imposti i vincoli:

$$\lambda_{irs}^{AVW} = \lambda_{jrs}^{BVW} = \lambda_{krs}^{CVW} = \lambda_{lrs}^{DVW} = 0$$

Il modello a classi latenti *basic* con R fattori latenti dicotomici ha lo stesso numero di parametri distinti di un modello a classi latenti tradizionale con $R+1$ classi. La parametrizzazione a classi latenti fattoriale permette di specificare un modello con 2^R classi con lo stesso numero di parametri di un modello a classi latenti tradizionale con $R+1$ classi. Questo porta ad un gran vantaggio in termini di parsimonia rispetto ai modelli a classi latenti tradizionali con T classi poiché il numero di parametri è notevolmente ridotto dalle naturali restrizioni. Il modello fattoriale a classi latenti con R fattori dicotomici è perciò spesso identificato quando il modello tradizionale con 2^R classi non lo è.

4.3 I modelli di regressione a classi latenti

In quest'ultimo paragrafo di questo capitolo viene introdotto il modello di regressione a classi latenti. Esso può essere espresso tramite la seguente struttura di probabilità generale:

$$f(y_i | z_i^{cov}, z_i^{pred}) = \sum_{x=1}^K P(x | z_i^{cov}) \prod_{t=1}^{T_i} f(y_{it} | x, z_{it}^{pred})$$

dove:

- y_{it} è il valore della variabile dipendente osservata sull'unità i nella replicazione t ;
- T_i è il numero di osservazioni ripetute della singola variabile dipendente y_{it} ;
- z_i^{cov} è un vettore di covariate;
- z_i^{pred} è un vettore di predittori;
- x è una variabile latente nominale con K categorie, o classi.

La principale differenza tra i modelli a classi latenti cluster e fattoriali e il modello di regressione a classi latenti è che in quest'ultimo è presente una variabile dipendente, la quale può essere osservata più volte (T_i) per ciascun soggetto i . Nel modello di regressione a classi latenti sono presenti due tipi di variabili esogene: variabili che influenzano la variabile latente X , chiamate covariate, e variabili che influenzano la variabile dipendente y_i , chiamate predittori. Inoltre in questo modello viene assunto che le densità di probabilità condizionata

$f(y_{it}|x, z_{it}^{pred})$ abbiano la stessa forma per ogni t e non sono presenti effetti diretti tra le risposte multiple.

La struttura di probabilità più semplice per un modello di regressione a classi latenti si ha quando c'è un'unica risposta per individuo, ovvero $T_i=1$ per ogni i , e non sono presenti predittori. Esso viene espresso nel modo seguente:

$$f(y_i) = \sum_{x=1}^K P(x) f(y_i|x)$$

Si tratta di un semplice modello misto, finito e univariato in cui la media ed eventualmente anche la varianza della distribuzione di y_i risultano dipendere dalla classe considerata. Un modello di questo tipo permette di descrivere l'eterogeneità non osservata riguardo alla distribuzione di y_i .

Risulta più utile un modello di regressione a classi latenti che include dei predittori, come per esempio:

$$f(y_i|z_{i1}^{pred}, z_{i2}^{pred}) = \sum_{x=1}^K P(x) f(y_i|x, z_{i1}^{pred}, z_{i2}^{pred})$$

dove:

- $f(y_i|X, z_{i1}^{pred}, z_{i2}^{pred})$ indica la distribuzione della variabile dipendente y data l'appartenenza ad una classe latente x e dati i valori dei predittori z_{i1}^{pred} e z_{i2}^{pred} .

Un'estensione di questo modello si ottiene introducendo delle covariate da cui dipende l'appartenenza ad una certa variabile latente. Un esempio di questo modello è formulato di seguito:

$$f(y_i|z_{i1}^{cov}, z_{i2}^{cov}, z_{i1}^{pred}, z_{i2}^{pred}) = \sum_{x=1}^K P(x|z_{i1}^{cov}, z_{i2}^{cov}) f(y_i|x, z_{i1}^{pred}, z_{i2}^{pred}) \quad (24)$$

In questo modello la probabilità di appartenere ad una certa classe latente x dipende dai valori assunti dalle covariate z_{i1}^{cov} e z_{i2}^{cov} .

Come già visto in precedenza ci può essere più di una osservazione per ciascun caso, ovvero potrebbero esserci più ripetizioni per la variabile dipendente e per le variabili indipendenti, per ogni unità osservata. Estendendo il modello (24) tramite l'introduzione di osservazione replicate si ottiene la seguente struttura di probabilità:

$$f(\mathbf{y}_i | \mathbf{z}_{i1}^{cov}, \mathbf{z}_{i2}^{cov}, \mathbf{z}_{i1}^{pred}, \mathbf{z}_{i2}^{pred}) = \sum_{x=1}^K P(x | \mathbf{z}_{i1}^{cov}, \mathbf{z}_{i2}^{cov}) \prod_{t=1}^{T_i} f(\mathbf{y}_i | x, \mathbf{z}_{i1}^{pred}, \mathbf{z}_{i2}^{pred})$$

CAPITOLO 5

VALIDAZIONE DEL QUESTIONARIO CON L'UTILIZZO DI MODELLI A CLASSI LATENTI

In questo capitolo alcuni dei modelli a classi latenti visti nel capitolo precedente vengono applicati ai dati provenienti dai questionari di soddisfazione della didattica. In particolare viene valutata la dimensionalità della scala e la validità del criterio.

Nel valutare la dimensionalità della scala l'analisi fattoriale viene sostituita da un *modello a classi latenti fattoriale*. Per la valutazione della validità del criterio gli indici di correlazione e l'ANOVA (che richiede una distribuzione normale) vengono sostituiti da un *modello a classi latenti di tipo cluster*, che tiene conto della natura non osservabile dell'oggetto d'interesse. Inoltre, al posto dell'analisi della correlazione, si può utilizzare un *modello di regressione a classi latenti* per la validazione del costrutto, questa proprietà non può però essere verificata con i dati di questa tesi per mancanza di item di validazione.

Lo scopo di questo capitolo consiste nel valutare se i modelli a classi latenti, i quali considerano esplicitamente la natura ordinale delle variabili e assumono che il costrutto d'interesse non sia direttamente osservabile, possano risultare più adeguati nella validazione di una scala di misura rispetto alle tecniche statistiche più tradizionali viste nei Capitoli 1 e 3.¹⁰

¹⁰ I risultati riportati in questo capitolo sono stati ottenuti tramite l'utilizzo del software Latent GOLD 5.0 (Vermunt e Magidson, 2013).

5.1 Dimensionalità della scala

Nel Capitolo 3 per determinare la dimensionalità della scala viene utilizzata l'analisi fattoriale. Sebbene questa tecnica sia adatta nel caso in cui le variabili siano di natura quantitativa, essa viene utilizzata anche quando le variabili sono di tipo ordinale. In questo caso, però, i risultati possono essere distorti e gli indici di adattamento del modello possono risultare errati (Bassi, 2011). L'utilizzo di un modello fattoriale a classi latenti permette di trattare le variabili generate dagli item come variabili nominali o ordinali.

Si ricorda che l'analisi fattoriale precedentemente condotta sui 17 item del questionario (è stato escluso l'item 14 relativo alla soddisfazione generale e considerato come *golden standard*) ha portato a due risultati:

- tramite la *Regola degli autovalori maggiori di 1* sono stati estratti **3 fattori**, i quali spiegano complessivamente il 71,29% della varianza; in particolare il **primo fattore** ne spiega il 56,87% e sembra descrivere la soddisfazione dell'attività didattica. Esso è caratterizzato dagli item **1, 2, 3, 8** (Organizzazione Didattica), **6, 7, 9** (Azione Didattica), **4, 10, 11, 15** e **5**, quest'ultimo non sembra però appartenere in maniera rilevante a questo fattore. Al **secondo fattore** sono legati gli item **16, 17** e **18**, i quali indagano aspetti che sembrano riferirsi alle preferenze del singolo studente e all'opinione sul corso messo in relazione al corso di studi e all'ambito lavorativo. Il **terzo fattore** è caratterizzato dagli item **12** e **13**, che sono legati ad aspetti di tipo logistico quali l'adeguatezza delle aule e dei laboratori.
- Aggiungendo un fattore ai tre precedenti risultano **4 fattori** in grado di spiegare il 75,67% della varianza; il **quarto componente**, dunque, contribuisce con una percentuale del 4,38% ed è caratterizzato dagli item **5** e **15**, connessi alle preconoscenze dello studente e al carico di studio dell'attività didattica.

5.1.1 Determinazione del numero di fattori

Sono stati stimati alcuni modelli a classi latenti fattoriali con diverso numero di fattori e diverso numero di livelli per fattore. Per effettuare un confronto tra questi modelli si sono osservati i valori riportati in Tabella 29: il valore della log-verosimiglianza (LL), il valore

dell'indice BIC basato sulla log-verosimiglianza (BIC_{LL}), il numero di parametri del modello (N° par.), il valore della statistica L^2 , i gradi di libertà della statistica L^2 (df), quando essa si distribuisce approssimativamente come una χ^2 , il p-value associato alla statistica L^2 e l'errore di classificazione. In particolare in Tabella 29 sono riportati i modelli in cui sono stati osservati i valori delle statistiche BIC_{LL} e L^2 più bassi. Basandosi su queste due grandezze i modelli con 4 fattori sembrano essere quelli che si adattano meglio ai dati, ma l'analisi delle loro caratteristiche porta ad escluderli: i coefficienti fattoriali non permettono una chiara identificazione dei fattori, l'ampiezza di alcuni livelli risulta troppo esigua e, inoltre, l'errore di classificazione risulta maggiore in questi modelli rispetto ai modelli a 3 fattori. Il modello che sembra adattarsi più adeguatamente ai dati è il modello a classi latenti con 3 fattori caratterizzati da 4 livelli ciascuno. Esso, infatti, oltre a presentare valori delle statistiche BIC_{LL} e L^2 inferiori rispetto al modello a 3 fattori con 3 livelli ciascuno, presenta livelli di ampiezza non troppo esigua (tutti i 4 livelli dei 3 fattori hanno ampiezza > 5%).

Modello	LL	BIC_{LL}	N° par.	L^2	df	p-value	Err. Class.
3 fattori (3,3,3)	-1.275.494,20	2.553.280,42	210	1.560.045,72	54.753	<0,001	0,035
3 fattori (4,4,4)	-1.256.502,67	2.515.330,10	213	1.522.062,66	54.750	<0,001	0,069
4 fattori (3,3,3,3)	-1.253.921,37	2.510.342,14	229	1.516.900,06	54.734	<0,001	0,083
4 fattori (4,4,4,4)	-1.235.849,26	2.474.241,59	233	1.480.755,85	54.730	<0,001	0,114

Tabella 29: Confronto tra 4 diversi modelli fattoriali a classi latenti.

Il modello a classi latenti fattoriale sembra, dunque, confermare che il costrutto d'interesse sia multidimensionale, in particolare ad esso sono associate 3 dimensioni. In Tabella 30 sono riportate le stime dei coefficienti fattoriali del modello a classi latenti con 3 fattori e 4 livelli per fattore, le quali consentono di individuare a quale delle 3 dimensioni ogni item è associato. Si nota che i risultati ottenuti in precedenza, tramite l'analisi fattoriale, sono confermati:

- al **primo fattore** sono associati gli item **1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 e 15** che descrivono la soddisfazione nei confronti dell'attività didattica;
- al **secondo fattore** sono associati gli item **16, 17 e 18**, che indagano aspetti relativi al contenuto del corso;
- al **terzo fattore** sono associati gli item **12 e 13**, legati ad aspetti di tipo logistico.

Item	Fattore 1	Fattore 2	Fattore 3	R ²
F01	-0,5981	-0,4840	-0,1877	0,8054
F02	-0,5851	-0,4562	-0,1997	0,7617
F03	-0,5430	-0,4570	-0,2003	0,7052
F04	-0,5630	-0,4424	-0,2645	0,7185
F05	-0,5084	-0,4565	-0,2023	0,5745
F06	-0,6257	-0,5102	-0,1465	0,8431
F07	-0,6454	-0,4804	-0,1587	0,8564
F08	-0,6231	-0,4710	-0,1986	0,8000
F09	-0,5896	-0,4834	-0,1969	0,8201
F10	-0,5717	-0,4821	-0,2120	0,7962
F11	-0,5638	-0,4854	-0,2839	0,7864
F12	-0,2654	-0,4553	-0,6882	0,8095
F13	-0,2981	-0,4733	-0,6901	0,8577
F15	-0,5298	-0,4850	-0,2309	0,6689
F16	-0,4763	-0,6754	-0,0616	0,7883
F17	-0,4104	-0,7183	-0,0451	0,7784
F18	-0,3990	-0,7001	-0,0637	0,7291

Tabella 30: Stime dei coefficienti fattoriali del modello a classi latenti con 3 fattori e 4 livelli ciascuno.

Per valutare l'affidabilità di ciascun fattore si calcola il coefficiente *Alpha di Cronbach* per ognuna delle tre dimensioni identificate. I valori dei coefficienti risultano pari a 0,967 per il primo fattore, 0,899 per il secondo fattore e 0,888 per il terzo fattore, segnalando un alto grado di affidabilità interna.

Si vuole sottolineare il fatto che il coefficiente Alpha di Cronbach, che si basa sulla varianza degli item, non è adatto a variabili di tipo ordinale. In alternativa è possibile utilizzare un indice chiamato "*ordinal alpha*"¹¹ che, invece di basarsi sulla matrice di correlazione di Pearson, si basa sulla matrice di correlazione policorica¹² e che, quindi, stima in maniera più precisa l'indice Alpha nel caso di variabili di natura ordinale. Tuttavia, l'accuratezza dell'Alpha di Cronbach sembra aumentare al crescere dei punteggi di risposta. Nel caso del questionario oggetto di studio, caratterizzato da 10 punteggi di risposta per item, i risultati ottenuti dal calcolo dell'Alpha di Cronbach si possono ritenere abbastanza accurati¹³.

¹¹ (Gadermann, Guhn, & Zumbo, 2012)

¹² Il coefficiente di correlazione policorica è una misura di associazione per variabili ordinali che si basa sull'assunzione di una distribuzione continua sottostante.

¹³ Si è provato a calcolare l'*ordinal alpha* con il software statistico R, ma non è stato possibile calcolare le correlazioni policoriche, in quanto in presenza di un numero di categorie maggiore di 8 tali correlazioni non sono probabilmente necessarie.

5.1.2 Descrizione dei livelli identificati

Dopo aver identificato e caratterizzato i fattori latenti sottostanti alla scala di misura, si passa all'analisi e alla caratterizzazione dei 4 livelli di ciascun fattore. A questo scopo si osservano le probabilità condizionate di assegnare la risposta i (con $i= 1, 2, \dots, 10$) ad un determinato item, data l'appartenenza ad uno dei 3 fattori e ad uno dei 4 livelli di ciascun fattore; queste probabilità sono riportate nelle tabelle 31, 32 e 33, illustrate di seguito.

In particolare, in Tabella 31 sono riportati i valori dell'ampiezza di ogni livello del **primo fattore** (in ordine crescente: 6,68%, 12,35%, 13,15% e 67,81%) e le probabilità condizionate di due dei 12 item collegati al primo fattore (per semplicità si riportano solo le probabilità condizionate relative agli item 2 e 9); considerando i risultati ottenuti anche negli altri 10 item si può affermare che i 4 livelli del primo fattore si caratterizzano nel modo seguente:

- nel primo livello, di ampiezza pari a **13,15%**, i soggetti tendono ad assegnare alle risposte punteggi non inferiori a 8, in particolare il punteggio pari a **10** è quello che presenta la probabilità più alta in tutti i 12 item associati al primo fattore. Le probabilità in corrispondenza dei punteggi 8 e 9, seppur non nulle, risultano basse;
- nel secondo livello, di ampiezza pari a **12,35%**, le probabilità di assegnare un punteggio inferiore a 7 sono nulle o quasi nulle; anche in questo caso la probabilità maggiore si trova in corrispondenza del punteggio **10**. Rispetto al primo livello aumentano le probabilità condizionate dei punteggi 8 e 9;
- nel terzo livello, di ampiezza pari a **67,81%**, aumentano in minima parte le probabilità di assegnare punteggi bassi, come 1, 2, 3, 4, 5 e 6, rimanendo comunque molto basse. La probabilità in corrispondenza del punteggio 10 risulta molto minore rispetto ai due livelli precedenti e le probabilità maggiori si hanno in corrispondenza dei punteggi pari a 7, 8 e 9 (in 8 si ha la probabilità maggiore);
- nel quarto ed ultimo livello, di ampiezza pari a **6,68%**, la maggior probabilità risulta in corrispondenza del punteggio pari a **1**, a seguire i punteggi **6** e **7**. Rimangono sempre molto basse le probabilità di osservare un punteggio pari a 2, 3, 4 o 5.

		Fattore 1			
		Livello 1	Livello 2	Livello 3	Livello 4
		13,15%	12,35%	67,81%	6,68%
Item					
F02	1	0,000	0,000	0,007	0,291
	2	0,000	0,000	0,006	0,087
	3	0,000	0,000	0,009	0,069
	4	0,000	0,000	0,014	0,059
	5	0,000	0,000	0,034	0,087
	6	0,000	0,004	0,102	0,156
	7	0,000	0,025	0,185	0,148
	8	0,007	0,112	0,281	0,083
	9	0,091	0,306	0,254	0,018
	10	0,903	0,553	0,107	0,001
F09	1	0,000	0,000	0,009	0,346
	2	0,000	0,000	0,006	0,075
	3	0,000	0,000	0,008	0,055
	4	0,000	0,000	0,012	0,047
	5	0,000	0,000	0,032	0,079
	6	0,000	0,002	0,099	0,157
	7	0,000	0,017	0,182	0,155
	8	0,002	0,088	0,275	0,075
	9	0,037	0,234	0,247	0,011
	10	0,961	0,659	0,131	0,001

Tabella 31: Probabilità condizionate di rispondere *i* all'item 2 e all'item 9 data l'appartenenza ad uno dei 4 livelli del fattore 1.

In Tabella 32 sono riportati i valori dell'ampiezza di ogni livello del **secondo fattore** (in ordine crescente: 15,67%, 22,90%, 29,47% e 31,96%) e le probabilità condizionate di tutti i 3 item associati al fattore: item 16, 17 e 18. I 4 livelli del secondo fattore si caratterizzano nel modo seguente:

- nel primo livello, di ampiezza pari a **31,96%**, la più alta probabilità si osserva in corrispondenza del punteggio pari a **10**, a seguire il punteggio pari a **9** presenta la seconda probabilità più alta. Gli altri punteggi presentano, invece, probabilità molto basse, in particolare i punteggi minori di 8 presentano probabilità pari a zero o prossime a zero;
- nel secondo livello, di ampiezza pari a **29,47%**, diminuisce rispetto al livello precedente la probabilità di assegnare un punteggio pari a 10 agli item. Aumentano, invece, le probabilità di scegliere i punteggi 7 e 8, in particolare in 8 si osserva la

probabilità maggiore. Si nota un leggero aumento delle probabilità in corrispondenza dei punteggi 5 e 6, che rimangono comunque molto basse. In generale i rispondenti che appartengono a questo livello tendono a scegliere punteggi pari a **7, 8, 9 o 10**;

- nel terzo livello, di ampiezza pari a **22,90%**, la probabilità maggiore si ha in corrispondenza del punteggio pari a 7. Diminuisce ulteriormente la probabilità di scegliere il punteggio pari a 10, mentre aumenta quella di scegliere un punteggio pari a 6. In generale, data l'appartenenza a questo livello, gli studenti tendono ad assegnare punteggi pari a **6, 7, 8 o 9**;
- nel quarto e ultimo livello, di ampiezza pari a **15,67%**, aumentano le probabilità di assegnare il punteggio pari a **1 o 5**, rispetto ai 3 livelli precedenti; la probabilità più alta si osserva in corrispondenza del punteggio pari a **6**. Le probabilità di scegliere i punteggi 8, 9 e 10 risultano molto basse.

		Fattore 2			
		Livello 1	Livello 2	Livello 3	Livello 4
		31,96%	29,47%	22,90%	15,67%
Item					
F16	1	0,000	0,000	0,013	0,101
	2	0,000	0,000	0,008	0,038
	3	0,000	0,000	0,008	0,049
	4	0,000	0,000	0,009	0,066
	5	0,000	0,003	0,026	0,146
	6	0,002	0,028	0,160	0,265
	7	0,020	0,137	0,345	0,156
	8	0,123	0,363	0,234	0,099
	9	0,342	0,258	0,107	0,064
	10	0,513	0,211	0,090	0,017
F17	1	0,000	0,000	0,004	0,109
	2	0,000	0,000	0,003	0,037
	3	0,000	0,000	0,004	0,046
	4	0,000	0,000	0,007	0,063
	5	0,000	0,001	0,024	0,139
	6	0,000	0,016	0,151	0,273
	7	0,009	0,108	0,349	0,174
	8	0,082	0,341	0,267	0,106
	9	0,309	0,307	0,120	0,047
	10	0,599	0,226	0,073	0,008

F18	1	0,000	0,000	0,005	0,126
	2	0,000	0,000	0,003	0,048
	3	0,000	0,000	0,005	0,056
	4	0,000	0,000	0,011	0,077
	5	0,000	0,003	0,043	0,143
	6	0,002	0,036	0,195	0,241
	7	0,025	0,164	0,325	0,170
	8	0,148	0,343	0,247	0,099
	9	0,333	0,266	0,111	0,035
	10	0,492	0,189	0,054	0,007

Tabella 32: Probabilità condizionate di rispondere *i* agli item 16, 17 e 18 data l'appartenenza ad uno dei 4 livelli del fattore 2.

In Tabella 33 sono riportati i valori dell'ampiezza di ogni livello del **terzo fattore** (in ordine crescente: 5,33%, 13,85%, 14,60% e 66,23%) e le probabilità condizionate dei 2 item associati al fattore: item 12 e 13. I 4 livelli del terzo fattore si caratterizzano nel modo seguente:

- nel primo livello, di ampiezza pari a **14,60%**, si hanno probabilità nulle o quasi nulle in corrispondenza dei punteggi minori di 8. Le più alte probabilità sono associate ai punteggi pari a **8, 9 e 10** (in 10 si osserva la probabilità maggiore);
- nel secondo livello, di ampiezza pari a **66,23%**, le probabilità più alte si osservano in corrispondenza dei punteggi **7, 8, 9 e 10** (in 8 si osserva la probabilità maggiore). Mentre rimangono molto basse le probabilità di osservare un punteggio pari a 1, 2, 3, 4 o 5;
- nel terzo livello, di ampiezza pari a **13,85%**, aumentano le probabilità in corrispondenza dei punteggi pari a **5, 6 e 7** (in 6 si osserva la probabilità massima). Le probabilità di assegnare il punteggio 9 o 10 risultano molto basse;
- nel quarto ed ultimo livello, di ampiezza pari a **5,33%**, la più alta probabilità si nota in corrispondenza del punteggio pari a **1** e aumentano leggermente le probabilità in corrispondenza dei punteggi pari a **2, 3 e 4**. Le probabilità di assegnare un punteggio maggiore di 5 risultano invece molto basse.

Fattore 3					
		Livello 1	Livello 2	Livello 3	Livello 4
		14,60%	66,23%	13,85%	5,33%
Item					
F12	1	0,000	0,000	0,064	0,415
	2	0,000	0,001	0,046	0,133
	3	0,000	0,003	0,063	0,117
	4	0,000	0,010	0,081	0,102
	5	0,001	0,036	0,132	0,101
	6	0,007	0,105	0,214	0,083
	7	0,038	0,187	0,209	0,038
	8	0,133	0,277	0,138	0,011
	9	0,248	0,221	0,042	0,001
	10	0,575	0,161	0,011	0,000
F13	1	0,000	0,000	0,066	0,421
	2	0,000	0,001	0,045	0,118
	3	0,000	0,002	0,056	0,102
	4	0,000	0,008	0,073	0,100
	5	0,000	0,031	0,120	0,107
	6	0,004	0,101	0,215	0,092
	7	0,032	0,181	0,222	0,044
	8	0,125	0,275	0,144	0,014
	9	0,250	0,246	0,048	0,001
	10	0,588	0,157	0,011	0,000

Tabella 33: Probabilità condizionate di rispondere *i* agli item 12 e 13 data l'appartenenza ad uno dei 4 livelli del fattore 3.

Utilizzando i modelli a classi latenti fattoriali, dunque, viene confermata la natura multidimensionale del costrutto d'interesse, indentificata anche grazie all'analisi fattoriale. In particolare vengono confermati i 3 fattori individuati con l'utilizzo dell'approccio tradizionale, tramite la *Regola deli autovalori maggiori di 1*; viene, invece, esclusa la presenza di un quarto fattore caratterizzato dagli item 5 e 15.

L'analisi a classi latenti, rispetto all'analisi fattoriale, ha permesso un'ulteriore caratterizzazione dei tre fattori. È stato possibile, infatti, identificare e descrivere i 4 livelli da cui ogni fattore è caratterizzato. I risultati ottenuti hanno permesso di fare alcune considerazioni sulla scala di misura studiata.

- Nei 4 livelli di ogni fattore si nota che le probabilità di assegnare agli item un punteggio pari a 1, 2, 3, 4 o 5 risultano in generale molto basse e spesso nulle; inoltre, i livelli in cui gli studenti tendono a scegliere questi punteggi sono quelli con ampiezza più contenuta. Ciò significa che i rispondenti tendono a non utilizzare spesso queste categorie per esprimere il loro grado di soddisfazione nei confronti della didattica.
- Risulta interessante il fatto che i 4 livelli dei fattori richiama la scala a 4 passi di cui si è parlato nel paragrafo 3.4. Le probabilità riportate nelle tabelle 31, 32 e 33 sembrano, infatti, dividere i rispondenti in 4 categorie. Per identificare e descrivere queste categorie risulta utile la Tabella 34, in cui si osservano i valori medi di soddisfazione in corrispondenza di ciascun livello dei 3 fattori; le medie sono state calcolate considerando solo gli item appartenenti al fattore corrispondente (per esempio le medie del secondo fattore sono state calcolate sugli item 16, 17 e 18, ecc.). I valori medi nei primi due fattori risultano abbastanza simili in tutti i 4 livelli: nel primo livello la media risulta molto alta e vicina al valore 10, nel secondo livello il valore scende leggermente e si aggira attorno a 9, nel terzo livello assume un valore compreso tra 7 e 8 e, infine, nell'ultimo livello è pari a 4,15 per il primo fattore e 5,41 per il secondo. I valori medi osservati nel terzo fattore si differenziano rispetto ai precedenti; essi, infatti, tendono ad essere più bassi, soprattutto in corrispondenza dei livelli 3 e 4 dove assumono valore pari a 5,71 e 2,84 rispettivamente.

Si può, inoltre, osservare che i valori medi di soddisfazione nei quattro livelli di ciascun fattore non risultano equidistanti tra di loro: le distanze tra i primi due livelli, per esempio, risultano tutte più piccole rispetto alle distanze tra gli ultimi due livelli. Ciò sottolinea l'importanza di considerare le variabili come ordinali, anziché quantitative, tenendo conto del fatto che le distanze tra punteggi adiacenti potrebbero non risultare uguali.

Fattore 1				
	Livello 1	Livello 2	Livello 3	Livello 4
Ampiezza classi	13,15%	12,35%	67,81%	6,68%
Media complessiva*	9,82	9,23	7,60	4,15
Fattore 2				
	Livello 1	Livello 2	Livello 3	Livello 4
Ampiezza classi	31,96%	29,47%	22,90%	15,67%
Media complessiva*	9,38	8,50	7,34	5,41
Fattore 3				
	Livello 1	Livello 2	Livello 3	Livello 4
Ampiezza classi	14,60%	66,23%	13,85%	5,33%
Media complessiva*	9,36	8,01	5,71	2,84

*Calcolata sugli item appartenenti al fattore

Tabella 34: Livello medio di soddisfazione nei 4 livelli dei tre fattori.

Queste osservazioni portano a riflettere sul possibile sviluppo di una scala di misura con un numero di punteggi di risposta minore. I rispondenti, infatti, non sembrano sfruttare tutte le 10 categorie di risposta, ma sembrano concentrarsi principalmente sui punteggi più alti. Potrebbe risultare interessante l'utilizzo di una scala a 4 passi, o a 5 passi, nel caso si intenda inserire la possibilità di assegnare ai quesiti una risposta neutrale. Una scala di misura con un numero di passi minore potrebbe aiutare a far emergere in modo più evidente il livello di soddisfazione da parte degli studenti: per esempio, in una scala a 4 passi i punteggi pari a 1 segnalerebbero chiaramente un livello di soddisfazione molto basso, così come punteggi pari a 4 indicherebbero un livello di soddisfazione molto alto.

La scelta di un punteggio da 1 a 10 potrebbe condurre il rispondente, per esempio, ad assegnare indifferentemente un punteggio pari a 8, 9 o 10; in una scala a 4 passi la distanza tra i punteggi potrebbe essere percepita come più ampia, risultando in una scelta più precisa e attenta: la scelta tra 3 o 4 risulta più determinante della scelta tra 9 o 10.

Come detto sopra, dall'osservazione delle probabilità condizionate si nota che i punteggi bassi vengono utilizzati poco o raramente, ciò potrebbe essere dovuto ad un effettivo medio-alto grado di soddisfazione da parte degli studenti oppure al fatto che di fronte ad un'ampia scelta di punteggi i rispondenti manifestino il loro grado di insoddisfazione assegnando agli item

punteggi medi (come 5 o 6). In entrambi i casi una scala di misura a 4 passi potrebbe rivelarsi utile, segnalando in maniera evidente i casi di molta insoddisfazione (pari a 1) e insoddisfazione (pari a 2), dai casi di soddisfazione (pari a 3): studenti che in una scala da 1 a 10 assegnano punteggi medi, in una scala a 4 passi, per esempio, sceglierebbero presumibilmente di assegnare un punteggio pari a 3 se il loro livello di soddisfazione fosse effettivamente medio, oppure la loro scelta ricadrebbe tra 1 o 2 se il loro livello di soddisfazione fosse molto basso o basso.

5.2 Validità del criterio

In questo paragrafo si propone un approccio alternativo a quello visto nel paragrafo 3.2 per la verifica della validità del criterio. Nel nuovo approccio le analisi di correlazione e della varianza (ANOVA) vengono sostituite da modelli a classi latenti di tipo cluster, i quali considerano esplicitamente che l'oggetto misurato non sia direttamente osservabile e che gli item generino variabili di tipo ordinale.

Dal momento che viene confermata l'esistenza di tre dimensioni sottese al costrutto d'interesse, è possibile effettuare la verifica della validità del criterio per ognuna di esse.

5.2.1 Determinazione del numero di cluster

Per poter stabilire il numero di cluster da cui è costituito ciascun fattore si sono stimati alcuni modelli con diverso numero di classi latenti. Nelle tabelle 35, 37 e 39, illustrate di seguito, si osservano le stesse grandezze riportate in Tabella 29, grazie alle quali è possibile effettuare un confronto tra i modelli.

5.2.1.1 *Primo fattore*

In Tabella 35 sono riportati i risultati ottenuti dalla stima di 5 modelli a classi latenti di tipo cluster per i 12 item relativi al primo fattore (item 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 15). Secondo le statistiche BIC_{LL} e L^2 il modello migliore è quello con 5 classi latenti, ma esso presenta

l'errore di classificazione con valore maggiore tra quelli in tabella e l'ampiezza di una delle sue classi risulta troppo contenuta (pari al 4,76%). Inoltre, la statistica L^2 si riduce solo del 4,81% rispetto al modello a 4 classi, ciò indica che il passaggio da un modello con 4 classi ad uno con 5 non porta ad un miglioramento rilevante. Il modello preferibile è, dunque, quello con 4 classi, le quali assumono livelli di ampiezza pari a 15,17%, 16,36%, 29,75% e 38,72%.

Modello	LL	BIC_{LL}	N° par.	L^2	df	p-value	Err. Class.
1 classe	-1.339.315,00	2.679.818,42	108	1.636.724,54	59.998	<0,001	0,00
2 classi	-1.146.688,74	2.294.708,96	121	1.251.472,02	59.985	<0,001	0,014
3 classi	-1.071.651,29	2.144.777,09	134	1.101.397,11	59.972	<0,001	0,025
4 classi	-1.023.217,34	2.048.052,26	147	1.004.529,22	59.959	<0,001	0,033
5 classi	-999.073,87	1.999.908,35	160	956.242,27	59.946	<0,001	0,045

Tabella 35: Confronto tra 5 diversi modelli a classi latenti di tipo cluster (primo fattore).

In Tabella 36 sono riportate le stime dei parametri legati agli effetti di secondo ordine del modello con 4 classi latenti, i quali descrivono la relazione tra gli indicatori e le classi, accanto ad esse sono riportati i valori della statistica di Wald, usata per esaminare se le stime dei parametri in corrispondenza dello stesso item siano tutte uguali a zero, i p-value ad essa associati e nell'ultima colonna si osservano i valori dell'indice R^2 , che indica in che misura ogni item è spiegato dal modello.

Si osserva che il valore del p-value risulta prossimo a zero in corrispondenza di tutti gli item, indicando che tutti gli indicatori contribuiscono in modo significativo a discriminare tra le classi.

Item	Classe 1	Classe 2	Classe 3	Classe 4	Wald	p-value	R^2
F01	-1,414	0,383	-2,903	3,934	16842,21	<0,001	0,647
F02	-1,180	0,251	-2,347	3,275	16520,1	<0,001	0,602
F03	-1,102	0,093	-2,008	3,018	15920,34	<0,001	0,533
F04	-0,991	0,113	-1,857	2,736	16495,85	<0,001	0,554
F05	-0,554	0,105	-1,113	1,562	15684,78	<0,001	0,444
F06	-1,506	0,433	-3,019	4,092	16252,07	<0,001	0,689
F07	-1,699	0,450	-3,370	4,619	15685,26	<0,001	0,706
F08	-1,372	0,189	-2,595	3,779	16766,98	<0,001	0,649
F09	-1,439	0,296	-3,101	4,244	15378,8	<0,001	0,654
F10	-1,377	0,211	-2,769	3,936	15703,63	<0,001	0,617
F11	-1,166	0,209	-2,274	3,231	16663,43	<0,001	0,609
F15	-0,764	0,066	-1,418	2,117	15537,99	<0,001	0,508

Tabella 36: Stima degli effetti di secondo ordine del modello a classi latenti di tipo cluster con 4 gruppi (primo fattore).

5.2.1.2 Secondo fattore

In Tabella 37 sono riportati i risultati ottenuti dalla stima di 5 modelli a classi latenti di tipo cluster considerando i 3 item legati al secondo fattore (item 16, 17 e 18). Viene escluso il modello con 5 classi latenti perché presenta l'errore di classificazione con valore più alto tra quelli in tabella ed è caratterizzato da una classe con ampiezza troppo esigua (pari a 3,35%). I modelli che sembrano adattarsi più adeguatamente ai dati sono quelli con 3 e 4 classi latenti. Tra questi risulta preferibile il modello con 3 classi, in corrispondenza del quale si ha un errore di classificazione minore rispetto a quello a 4 classi e che presenta classi di ampiezza adeguata: in ordine crescente 22,58%, 27,52% e 49,90%.

Modello	LL	BIC_{LL}	N° par.	L^2	df	p-value	Err. Class.
1 classe	-818.323,43	1.636.968,24	27	337.348,13	972	<0.001	0,000
2 classi	-738.552,60	1.477.474,19	31	177.806,47	968	<0.001	0,050
3 classi	-708.283,23	1.416.983,07	35	117.267,73	964	<0.001	0,075
4 classi	-693.047,61	1.386.559,44	39	86.796,49	960	<0.001	0,094
5 classi	-687.195,98	1.374.903,81	43	75.093,24	956	<0.001	0,126

Tabella 37: Confronto tra 5 diversi modelli a classi latenti di tipo cluster (secondo fattore).

Si osserva che i valori dei p-value riportati in Tabella 38, relativi alla statistica di Wald, risultano tutti prossimi a zero; anche nel secondo fattore, dunque, tutti gli item contribuiscono a misurare il costrutto latente sotteso.

Item	Classe 1	Classe 2	Classe 3	Wald	p-value	R ²
F16	-0,267	-1,832	2,099	22.724,22	<0,001	0,559
F17	-0,566	-3,117	3,683	13.770,78	<0,001	0,632
F18	-0,333	-1,643	1,977	16.890,00	<0,001	0,559

Tabella 38: Stima degli effetti di secondo ordine del modello a classi latenti di tipo cluster con 3 gruppi (secondo fattore).

5.2.1.3 Terzo fattore

In Tabella 39 sono riportati i risultati ottenuti dalla stima di 5 modelli a classi latenti di tipo cluster considerando come indicatori i 2 item associati al terzo ed ultimo fattore (item 12 e 13). Secondo le statistiche BIC_{LL} e L^2 i modelli che si adattano meglio ai dati sono quelli con 4

e 5 classi latenti; essi, però, presentano i valori dell'errore di classificazione più elevati tra quelli osservati in tabella e alcune delle loro classi risultano di ampiezza troppo esigua. Il modello con 3 classi latenti, dunque, risulta preferibile ed è caratterizzato da gruppi di ampiezza pari a 16,01%, 35,76% e 48,23%.

Modello	LL	BIC_{LL}	N° par.	L^2	df	p-value	Err. Class.
1 classe	-394.314,21	788.835,62	18	151.822,71	81	<0,001	0,000
2 classi	-364.878,95	729.999,63	21	92.952,19	78	<0,001	0,050
3 classi	-351.142,15	702.560,58	24	65.478,60	75	<0,001	0,089
4 classi	-344.036,63	688.384,06	27	51.267,55	72	<0,001	0,094
5 classi	-341.961,49	684.268,33	30	47.117,28	69	<0,001	0,106

Tabella 39: Confronto tra 5 diversi modelli a classi latenti di tipo cluster (terzo fattore).

I valori dei p-value relativi alla statistica di Wald, riportati in Tabella 40, indicano che entrambi gli item legati al secondo fattore contribuiscono significativamente a discriminare tra le classi.

Item	Classe 1	Classe 2	Classe 3	Wald	p-value	R ²
F12	-0,701	-2,611	3,312	9.672,82	<0,001	0,602
F13	-0,929	-2,913	3,842	7.932,06	<0,001	0,599

Tabella 40: Stima degli effetti di secondo ordine del modello a classi latenti di tipo cluster con 3 gruppi (terzo fattore).

5.2.2 Descrizione dei cluster identificati

Dopo aver determinato il numero di cluster, da cui ciascun fattore è caratterizzato, è possibile descrivere i gruppi identificati. A tal scopo si osservano i valori dei livelli di ampiezza e dei livelli medi di soddisfazione riportati in Tabella 41. Si nota che questi valori sono simili a quelli riportati in Tabella 34, la differenza è che i valori della tabella sottostante derivano dalla stima di 3 modelli a classi latenti di tipo cluster, in cui ciascun modello considera indicatori diversi (il primo modello considera gli item 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 e 15; il secondo modello considera gli item 16, 17 e 18; il terzo modello considera gli item 12 e 13), mentre i valori riportati in Tabella 34 si riferiscono alla stima di un modello a classi latenti fattoriali caratterizzato da 3 dimensioni con 4 livelli ciascuna, in cui si sono considerati tutti i 17 item del questionario (viene escluso l'item 14).

Si osserva, dunque, che tutte le classi hanno dimensione maggiore al 15% e presentano livelli di soddisfazione diversi tra loro. In particolare, nelle classi meno ampie si osservano i livelli di soddisfazione più alti pari a 9,89, 9,80 e 9,93 per il primo, secondo e terzo fattore rispettivamente. La seconda classe in ordine di ampiezza è caratterizzata da medie pari a 5,09, 5,78 e 5,47, che rappresentano i valori medi più bassi per il primo, secondo e terzo fattore, rispettivamente. La classe più ampia del secondo e terzo fattore presenta un livello di soddisfazione pari a 8,34 e 8,23. Il primo fattore, al contrario degli altri due, è caratterizzato da 4 classi; tra queste le due più ampie sono caratterizzate da livelli medi di soddisfazione pari a 7,51 e 8,85.

Dall'analisi di questi risultati sembrano confermate alcune delle considerazioni fatte precedentemente: i livelli di soddisfazione medi sono tutti superiori a 5 indicando basse probabilità di assegnare punteggi inferiori a questo valore e, inoltre, non risultano equidistanti tra le classi.

Fattore 1				
	Classe 1	Classe 2	Classe 3	Classe 4
Ampiezza classi	38,72%	29,75%	16,36%	15,17%
Media complessiva	7,51	8,85	5,09	9,89
Fattore 2				
	Classe 1	Classe 2	Classe 3	
Ampiezza classi	49,90%	27,52%	22,58%	
Media complessiva	8,34	5,78	9,80	
Fattore 3				
	Classe 1	Classe 2	Classe 3	
Ampiezza classi	48,23%	35,76%	16,01%	
Media complessiva	8,23	5,47	9,93	

Tabella 41: Livelli di ampiezza e di soddisfazione media relativi alle classi dei 3 fattori identificati.

5.2.3 Verifica della validità del criterio utilizzando le variabili latenti

Dopo aver identificato i tre costrutti latenti e definito le classi ad essi associate si passa alla procedura di verifica della validità di criterio, che consiste nel confronto tra il costrutto latente e l'item 14 (quesito sulla soddisfazione generale dell'attività didattica), il quale viene considerato come variabile-criterio. Per verificare la presenza di associazione tra queste due variabili si conduce il test *Chi-quadrato di Pearson* e si calcola l'indice *gamma di Goodman e Kruskal*, un indice di cograduazione che misura l'associazione di due variabili di tipo ordinale e che varia tra -1 e 1 (il valore 0 indica l'assenza di associazione).

Per il primo fattore la statistica Chi-quadrato di Pearson risulta pari a 94.105,96, alla quale corrisponde un p-value pari a 0, che porta al rifiuto dell'indipendenza tra variabili. Anche il valore osservato dell'indice gamma, pari a **0,938** e con p-value nullo, evidenzia un forte grado di associazione tra la variabile latente e l'item 14. Si è giunti alle stesse conclusioni dall'osservazione delle statistiche Chi-quadrato calcolate per il secondo e terzo fattore, che risultano significative, e dall'indice gamma pari a **0,736** (p-value nullo) e **0,586** (p-value nullo) per il secondo e terzo fattore, rispettivamente. Da questi valori si può anche osservare che il primo fattore, legato alla soddisfazione nei confronti della didattica, è quello maggiormente associato al quesito sulla soddisfazione generale; anche il secondo fattore, legato ai contenuti del corso, presenta un livello di associazione abbastanza alto, mentre il terzo fattore, legato ad aspetti di tipo logistico, è quello che influisce in maniera minore, seppur significativa, sul livello di soddisfazione generale.

Secondo questi risultati, dunque, si può affermare che tutti i tre aspetti sottesi alla scala di misura godono della proprietà di validità di criterio, confermando i risultati ottenuti con l'utilizzo dell'approccio tradizionale. Le differenze tra l'approccio tradizionale e quello visto in questo paragrafo consiste nel sostituire le variabili MEDIA_17 (media aritmetica dei punteggi dei 17 item, escluso l'item 14), ASO (Aspetti Organizzativi, media aritmetica dei punteggi degli item 1, 2, 3 e 8) e AZD (Azione Didattica, media aritmetica dei punteggi degli item 6, 7 e 9) con le variabili latenti identificate grazie al modello a classi latenti di tipo cluster, e nell'utilizzare un indice di associazione adatto a variabili di tipo ordinale al posto dei coefficienti di correlazione.

5.2.4 Validità di criterio dell'intera scala

Un'ulteriore verifica della validità del criterio è stata condotta considerando l'intera scala di misura (escluso l'item 14). La stima di modelli a classi latenti di tipo cluster sull'insieme dei 17 item porta alla scelta del modello con 4 gruppi, che è risultato migliore in termini di errore di classificazione e di ampiezza dei gruppi. La statistica di Wald e il p-value ad essa associato portano ad affermare che tutti gli item contribuiscono in modo significativo a discriminare tra le classi. In Tabella 42 sono riportati i livelli di ampiezza e i livelli medi di soddisfazione dei 4 cluster identificati. Si osserva che tutte le classi hanno dimensione maggiore al 14% e presentano livelli di soddisfazione diversi tra loro. In particolare nella classe più ampia è incluso il 39,51% dei rispondenti e si osserva un livello di soddisfazione pari a 7,52, la seconda classe, di dimensione pari al 29,55%, è caratterizzata da un livello di soddisfazione poco maggiore rispetto al precedente, pari a 8,83. Le classi 3 e 4 rappresentano i gruppi con il minore e il maggiore livello di soddisfazione, rispettivamente. In particolare nella classe 3 si riscontra un livello di soddisfazione, pari a 5,23, molto differente rispetto agli altri tre gruppi. La classe 4, in cui si osserva un livello di soddisfazione pari a 9,87, è quella di dimensione più esigua (pari al 14,31%).

Ancora una volta si sono ottenuti risultati simili a quelli riscontrati nell'analisi del modello a classi latenti fattoriale e nell'analisi dei modelli a classi latenti di tipo cluster utilizzati per la verifica della validità di criterio dei tre fattori. Sembra, dunque, confermata la suddivisione dei rispondenti in quattro categorie con diverso livello di soddisfazione.

	Classe 1	Classe 2	Classe 3	Classe 4
Ampiezza classi	39,51%	29,55%	16,63%	14,31%
Media complessiva	7,52	8,83	5,23	9,87

Tabella 42: Ampiezza delle classi e livello medio di soddisfazione nei 4 cluster.

La statistica Chi-quadrato di Pearson risulta pari a 89.543,917, alla quale corrisponde un p-value pari a 0, che porta al rifiuto dell'indipendenza tra variabili. Anche il valore osservato dell'indice gamma, pari a 0,949 e con p-value nullo, evidenzia un forte grado di associazione tra la variabile latente e l'item 14. Secondo questi risultati, dunque, si può affermare che la scala di misura goda della proprietà di validità di criterio, confermando i risultati ottenuti con l'utilizzo dell'approccio tradizionale.

CONCLUSIONI

Lo scopo di questa tesi consisteva nel condurre la validazione di una scala di misura usando due approcci diversi: il primo prevede l'utilizzo di tecniche statistiche tradizionali, adatte a variabili di tipo quantitativo e non in grado di considerare la natura latente della variabile misurata, il secondo si basa sull'analisi a classi latenti, che, invece, prende esplicitamente in considerazione la natura ordinale o nominale delle variabili e il fatto che il costrutto d'interesse non sia direttamente osservabile.

I dati utilizzati in questa tesi provengono dall'indagine condotta online nell'Anno Accademico 2012/2013 sulla valutazione della didattica da parte degli studenti dell'Università degli Studi di Padova. I questionari di valutazione erano composti da 18 domande principali e gli studenti potevano assegnare ad esse un punteggio da 1 a 10, dove 10 indicava la situazione ottimale.

Per determinare la dimensionalità della scala l'analisi fattoriale è stata confrontata con l'utilizzo di modelli a classi latenti fattoriali, i quali sono adatti a variabili ordinali, tengono conto della natura latente del costrutto e non richiedono una distribuzione normale. In entrambi i casi la scala di misura è risultata multidimensionale, ma mentre dall'analisi fattoriale emerge che il numero di fattori sottesi al costrutto possano essere tre o quattro, l'approccio a classi latenti identifica chiaramente tre dimensioni: la prima è legata alla soddisfazione nei confronti della didattica (definizione di obiettivi e modalità d'esame, rispetto dell'orario di lezione, adeguatezza delle ore previste e delle conoscenze preliminari, stimolo verso la disciplina da parte del docente, chiarezza nell'esposizione degli argomenti, adeguatezza del materiale consigliato, disponibilità e reperibilità del docente, adeguatezza di esercitazioni/laboratori/seminari e del carico di studio), la seconda riguarda i contenuti del corso (interesse verso i contenuti del corso, coerenza rispetto agli obiettivi del corso di studi e adeguatezza in ambito lavorativo delle conoscenze fornite) e la terza descrive aspetti di tipo logistico (adeguatezza di aule e laboratori).

Per determinare la validità del criterio nell'approccio tradizionale sono stati utilizzati indici di correlazione e in quello a classi latenti si sono stimati modelli a classi latenti di tipo cluster. In entrambi i casi la proprietà di validità di criterio è risultata soddisfatta, ma il secondo tipo di approccio ha permesso di identificare la variabile latente e confrontarla con la variabile-

criterio (item 14 sulla soddisfazione generale) attraverso un indice di associazione adatto a variabili di tipo ordinale.

Inoltre, grazie all'analisi a classi latenti è stato possibile ricavare informazioni sui comportamenti dei soggetti intervistati. E', infatti, emerso che l'insieme di rispondenti sembra dividersi in 4 categorie, caratterizzate da diversi comportamenti di risposta; inoltre, si è osservato che la tendenza ad assegnare punteggi inferiori a 5 risulta in generale molto bassa. Da queste osservazioni può nascere una riflessione sul possibile sviluppo di una scala di misura con un numero di punteggi di risposta inferiore a 10 (per esempio una scala a 4 o 5 passi), che potrebbe evidenziare in maniera più rilevante i livelli di soddisfazione da parte degli studenti.

Nonostante i risultati in termini di affidabilità e validità della scala non siano stati differenti tra i due approcci, l'analisi a classi latenti ha fatto emergere delle caratteristiche importanti del questionario preso in considerazione, rivelandosi un metodo più preciso e più utile, rispetto ai protocolli tradizionali, nel processo di sviluppo e validazione di una scala di misura.

BIBLIOGRAFIA

- Bassi, F. (2011). Latent class analysis for marketing scale development. *International Journal of Market Research*, 53(2), pp. 209-230.
- Churchill, G. A. (1979, February). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research*, 16(1), pp. 64-73.
- CNVSU. (2010). *Progettazione, implementazione e validazione di un questionario per la valutazione della didattica erogata a studenti universitari*.
- De Vellis, R. F. (1991). *Scale Development. Theory and applications*. Newbury Park: Sage.
- Gadermann, A. M., Guhn, M., & Zumbo, B. D. (2012, January). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical assessment, research and evaluation*, 17(3), pp. 1-13.
- Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Measurement Theory for the Behavioral Sciences*. New York: Freeman.
- Goodman, L. A. (1974a). Explanatory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2), pp. 215-231.
- Goodman, L. A. (1974b, March). The analysis of systems of qualitative variables when some of the variables are unobservable. Part I: a modified latent structure approach. *American Journal of Sociology*, 79(5), pp. 1179-1259.
- Goodman, L. A. (1979, March). On the estimation of parameters in latent structure analysis. *Psychometrika*, 44(1), pp. 123-128.
- Guido, G. (1999). *Aspetti metodologici e operativi del processo di ricerca di marketing*. Padova: Cedam.
- Guido, G., Bassi, F., & Peluso, A. (2010). *La soddisfazione del consumatore. La misura della customer satisfaction nelle esperienze di consumo* (I ed.). Milano: Franco Angeli.
- Haberman, S. J. (1979). *Analysis of qualitative data* (Vol. 2). New York: Academic Press.
- Hair, J. F., Black, W. C., Babin, B., Anderson, R. E., & Tatham, R. E. (2006). *Multivariate Data Analysis* (6th ed.). New Jersey, NJ: Prentice Hall.
- Jacoby, J. (1978, April). Consumer Research: A State of the Art Review. *Journal of Marketing*, 42(2), pp. 87-96.
- Langeheine, R., Pannekoek, J., & Van de Pol, F. (1996, May). Bootstrapping Goodness-of-Fit Measures in Categorical Data Analysis. *Sociological methods & research*, 24(4), pp. 492-516.
- Lazarsfeld, P. F., & Henry, N. W. (1968). *Latent structure analysis*. Boston: Houghton Mifflin.
- Magidson, J., & Vermunt, J. (2004). Latent class models. In D. Kaplan, *The Sage Handbook of Quantitative Methodology for the Social Sciences* (pp. 175-198 (Chapter 10)). Thousand Oaks: Sage.

- Magidson, J., & Vermunt, J. K. (2001). Latent class factor and cluster models, bi-plots and related graphical displays. *Sociological Methodology*, 31(1), pp. 223-264.
- McCutcheon, A. L. (1987). *Latent Class Analysis*. Newbury Park: Sage.
- Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.
- Peter, J. P. (1979, February). Reliability: a review of psychometric basics and recent marketing practices. *Journal of Marketing Research*, 16(1), pp. 6-17.
- Research Unit N.3 (PRODIG project). (2015). *Tools for teaching assessment at Padua University: role, development and validation*. Technical Report Series N.1, Università degli Studi di Padova.
- Vermunt, J. K., & Magidson, J. (2002). Latent class cluster analysis. In J. A. Hagenaars, & A. L. McCutcheon, *Applied latent class analysis* (pp. 89-106). Cambridge UK: Cambridge University Press.
- Vermunt, J. K., & Magidson, J. (2005). *Latent GOLD 4.0 User's Guide*. Belmont, Massachusetts: Statistical Innovations Inc.
- Vermunt, J. K., & Magidson, J. (2013). *Technical Guide for Latent GOLD 5.0: Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations Inc.