

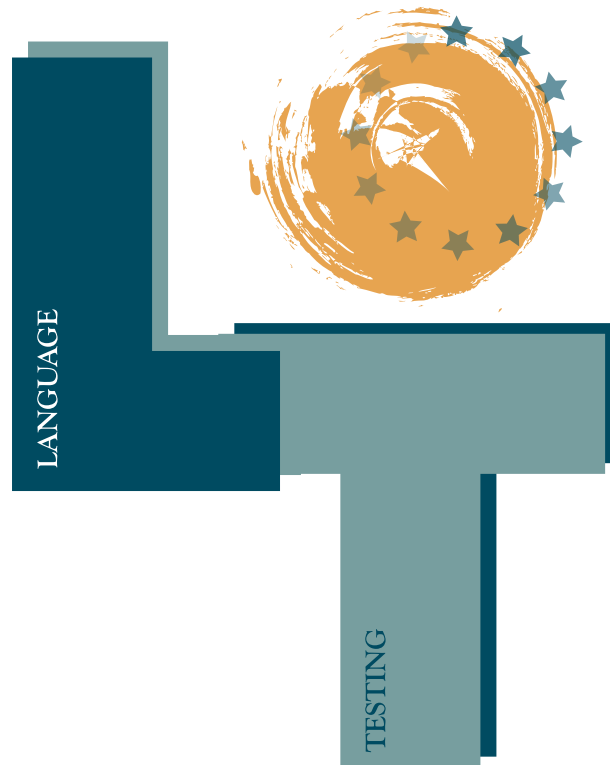
UNIVERSITÀ DEGLI STUDI DI PALERMO

Dipartimento di Scienze Filologiche e Linguistiche
Dottorato di Ricerca in
Linguistica Sincronica, Diacronica e Saperi Testuali
XXII ciclo

La valutazione linguistica

Sperimentazione di un modulo d'esame PFLISP

Settore Scientifico Disciplinare di afferenza L-LIN/02



Coordinatore:
Ch.mo prof. Lucio MELAZZO

Tesi di:
Simona LIBRIZZI

Tutor:
Ch.mo prof. Antonino DI SPARTI

Anni Accademici
2007/08 2008/09 2009/10

Indice	I
Indice delle figure	IV
Indice delle formule	V
Indice delle tabelle	VII
Abstract	1
Introduzione	2
Capitolo primo: Educazione plurilinguistica e certificazione	9
1.1 Premesse introduttive.....	9
1.2 Una realtà europea per le lingue.....	13
1.2.1 Panoramica storica.....	14
1.3 Il Quadro comune europeo di riferimento per le lingue.....	22
1.3.1 Obiettivi, criteri e approcci.....	24
1.3.2 La competenza comunicativa.....	28
1.3.3 Livelli comuni di riferimento.....	32
1.3.4 Il QCER: criteri, principi e tipologie di valutazione.....	36
1.3.5 Punti deboli del QCER.....	40
1.4 Il Portfolio Europeo delle Lingue.....	41
1.5 Valutazione.....	48
1.6 Certificazione.....	52
1.7 EALTA.....	54
1.8 ALTE.....	55
1.9 Riflessioni conclusive.....	61
Capitolo secondo: Testing linguistico e valutazione	63
2.1 Premesse introduttive.....	63
2.2 Presupposti teorici sui modelli di competenza linguistica e suggerimenti alla progettazione di <i>item</i>	65
2.3 Il modello di Bachman.....	71
2.4 L'utilità e le qualità essenziali di un test.....	80
2.4.1 Affidabilità.....	82
2.4.2 Validità di costrutto.....	84
2.4.3 Autenticità.....	86
2.4.4 Interattività.....	87
2.4.5 Impatto.....	89
2.4.6 Praticabilità.....	90
2.4.7 Osservazioni finali.....	92
2.5 Il modello VRIP del Cambridge ESOL.....	93
2.6 <i>Performance</i> e uso della lingua: le caratteristiche del compito.....	94
2.7 Le fasi di sviluppo del test.....	97
2.8 Fonti di errori nella valutazione.....	102
2.8.1 Validità e valutazione.....	104
2.8.2 Il modello di Weir.....	108
2.8.3 Affidabilità e valutazione.....	113
2.9 Tipologie di <i>item</i> e scale di valutazione.....	114
2.10 Riflessioni conclusive.....	120

Capitolo terzo: Statistica e analisi dei dati	123
3.1 Premesse introduttive	123
3.2 Le analisi statistiche	125
3.3 La statistica descrittiva	126
3.3.1 Analisi delle <i>performance</i>	137
3.3.2 Analisi degli <i>item</i>	157
3.3.3 Calcolo dell'affidabilità	176
3.3.4 Calcolo della validità	191
3.3.5 Interpretazione dei punteggi	193
3.4 Riflessioni conclusive	197
Capitolo quarto: Un caso di studio: PFLISP Sicilia	199
4.1 Premesse introduttive	199
4.2 Il Progetto PFLISP Sicilia	200
4.2.1 Fase I: elaborazione delle prove	206
4.2.2 Fase II: attuazione delle prove	212
4.2.3 Fase III: valutazione delle prove	213
4.3 Analisi dei dati del PFLISP Sicilia	214
4.3.1 Analisi statistica degli <i>item</i>	215
4.3.2 Analisi linguistica degli <i>item</i>	227
4.3.3 Analisi delle distribuzioni di punteggi	232
4.3.4 Analisi statistica delle <i>performance</i>	236
4.4 Riflessioni conclusive	240
Conclusione	242
Conclusioni generali	242
Conclusioni sul lavoro di analisi	243
Proposte di ricerca e applicazione della ricerca	244
Appendice A	246
Tabelle di riferimento estratte dal QCER	247
Griglia di autovalutazione	247
Domini	249
Tipologie di valutazione	250
Appendice B	251
Esempio di prova PFLISP Sicilia	251
Appendice C	259
Tabelle riassuntive dei dati raccolti	259
Gruppo A - Esiti prova di Listening e Reading n. 1	260
Gruppo <i>higher</i>	260
Gruppo <i>lower</i>	260
Deviazione standard	261
Punteggi Z e punteggi T	262
Differenze di posizione	264
Gruppo B - Esiti prova di Listening e Reading n. 2	265
Gruppo <i>higher</i>	266
Gruppo <i>lower</i>	266
Deviazione standard	267
Punteggi Z e punteggi T	268
Differenze di posizione	270
Gruppo C - Esiti prova di Listening e Reading n. 3	271
Gruppo <i>higher</i>	272
Gruppo <i>lower</i>	272
Deviazione standard	273
Punteggi Z e punteggi T	274

Differenze di posizione.....	276
Gruppo D - Esiti prova di Listening e Reading n. 4	277
Gruppo <i>higher</i>	278
Gruppo <i>lower</i>	278
Deviazione standard	279
Punteggi Z e punteggi T.....	280
Differenze di posizione.....	282
Gruppo E - Esiti prova di Listening e Reading n. 5	283
Gruppo <i>higher</i>	284
Gruppo <i>lower</i>	284
Deviazione standard	285
Punteggi Z e punteggi T.....	286
Differenze di posizione.....	287
Item Tally Sheet – Prova Listening n. 1	288
Item Tally Sheet – Prova Listening n. 2	289
Item Tally Sheet – Prova Listening n. 3	290
Item Tally Sheet – Prova Listening n. 4	291
Item Tally Sheet – Prova Listening n. 5	292
Item Tally Sheet – Prova Reading n. 1.....	293
Item Tally Sheet – Prova Reading n. 2.....	295
Item Tally Sheet – Prova Reading n. 3.....	297
Item Tally Sheet – Prova Reading n. 4.....	299
Item Tally Sheet – Prova Reading n. 5.....	301
Riferimenti bibliografici	303
Percorsi sitografici	323

INDICE DELLE FIGURE

Figura n. 1.1 - Il tradizionale modello di atto didattico	11
Figura n. 1.2 - La costellazione didattica	12
Figura n. 1.3 - Livelli comuni di riferimento	32
Figura n. 1.4 - P.E.L.	43
Figura n. 1.5 - Prima applicazione pratica	46
Figura n. 1.6 - Seconda applicazione pratica.....	47
Figura n. 1.7 - Rapporto tra <i>assessment</i> , misurazione, test	51
Figura n. 2.1 - Modello di Bachman	73
Figura n. 2.2 - Componenti della competenza linguistica	74
Figura n. 2.3 - Componenti dell'uso della lingua e della <i>performance</i>	76
Figura n. 2.4 - Corrispondenze tra l'uso della lingua e la <i>performance</i> del candidato.....	79
Figura n. 2.5 - Utilità	81
Figura n. 2.6 - Affidabilità	82
Figura n. 2.7 - Validità di costruito	85
Figura n. 2.8 - Autenticità	86
Figura n. 2.9 - Interattività	88
Figura n. 2.10 - Impatto.....	89
Figura n. 2.11- Praticabilità	91
Figura n. 2.12 - Componenti e qualità dell'utilità.....	92
Figura n. 2.13 - Il modello di Weir	110
Figura n. 3.1 - Curva di Gauss.....	129
Figura n. 3.2 - La curva di Gauss e la media	130
Figura n. 3.3 - Distribuzione dei punteggi in un grafico a istogramma	131
Figura n. 3.4 - Distribuzione dei punteggi in un poligono delle frequenze	132
Figura n. 3.5 - Diagramma a scatola.....	133
Figura n. 3.6 - Rappresentazione grafica dei quartili	133
Figura n. 3.7 - Rappresentazione della distribuzione in un <i>box-plot</i>	134
Figura n. 3.8 - Rappresentazione della mediana e dei quartili	135
Figura n. 3.9 - Poligono delle frequenze	142
Figura n. 3.10 - Diagramma a dispersione.....	149
Figura n. 3.11 - Linea di regressione di Y sulla base di X	155
Figura n. 3.12 - Rappresentazione della <i>Item Characteristic Curve</i> (ICC)	170
Figura n. 3.13 - Un esempio di rappresentazione della <i>Item Characteristic Curve</i>	172
Figura n. 3.14 - Esempio di ICC di due <i>item</i>	173
Figura n. 3.15 - Rappresentazione della <i>Item Information Function</i> (IIF)	174
Figura n. 3.16 - Esempio di ICC e IIF	174
Figura n. 3.17 - Varianza nel modello CTT	177
Figura n. 4.1 - Media dei punteggi di ciascun gruppo per le prove di ascolto	233
Figura n. 4.2 - Media dei punteggi di ciascun gruppo per le prove di lettura.....	234

INDICE DELLE FORMULE

Formula n. 2.1 Calcolo del punteggio con penalizzazione	116
Formula n. 3.1 Calcolo del valore adiacente inferiore	135
Formula n. 3.2 Calcolo del valore adiacente superiore	135
Formula n. 3.3 Calcolo della mediana	137
Formula n. 3.4 Calcolo della mediana 2	138
Formula n. 3.5 Calcolo della media aritmetica	138
Formula n. 3.6 Calcolo della media ponderata	139
Formula n. 3.7 Calcolo del campo di variazione	140
Formula n. 3.8 Calcolo della differenza interquartile	140
Formula n. 3.9 Calcolo della semidifferenza interquartile	140
Formula n. 3.10 Calcolo dello scarto medio	141
Formula n. 3.11 Calcolo dello scarto quadratico o . deviazione standard	141
Formula n. 3.12 Calcolo del punteggio Z	145
Formula n. 3.13 Calcolo del punteggio T	145
Formula n. 3.14 Calcolo della varianza	146
Formula n. 3.15 Calcolo della covarianza	151
Formula n. 3.16 Calcolo del coefficiente di Pearson	151
Formula n. 3.17 Calcolo del coefficiente di Pearson 2.	152
Formula n. 3.18 Calcolo del coefficiente di Spearman	153
Formula n. 3.19 Calcolo del coefficiente di regressione	154
Formula n. 3.20 Calcolo della pendenza della retta	155
Formula n. 3.21 Calcolo del segmento della retta	155
Formula n. 3.22 Calcolo dell'indice di difficoltà (prove R/W)	158
Formula n. 3.23 Calcolo dell'indice di difficoltà (percentuale)	158
Formula n. 3.24 Calcolo dell'indice di difficoltà (proporzione)	160
Formula n. 3.25 Calcolo dell'indice di difficoltà (gruppi h e l)	160
Formula n. 3.26 Calcolo dell'indice di distrattività (prove R/W; prove P-C)	161
Formula n. 3.27 Calcolo dell'indice di distrattività (gruppi h e l)	161
Formula n. 3.28 Calcolo dell'indice di distrattività (proporzione)	161
Formula n. 3.29 Calcolo dell'indice di difficoltà (prove P-C)	162
Formula n. 3.30 Calcolo dell'indice di difficoltà (gruppi h e l)	162
Formula n. 3.31 Calcolo dell'indice di difficoltà (punteggio Max e Min)	162
Formula n. 3.32 Calcolo dell'indice di discriminazione (prove R/W)	163
Formula n. 3.33 Calcolo dell'indice di discriminazione (proporzione)	163
Formula n. 3.34 Calcolo dell'indice di discriminazione del distrattore (gruppi h e l)	163
Formula n. 3.35 Calcolo dell'indice di discriminazione del distrattore (proporzione)	164
Formula n. 3.36 Calcolo dell'indice di discriminazione (prove P-C)	164
Formula n. 3.37 Calcolo dell'indice di distrattività (prove R/W)	165
Formula n. 3.38 Calcolo dell'indice di discriminazione (prove R/W)	165
Formula n. 3.39 Calcolo dell'indice di discriminazione (prove P-C)	166
Formula n. 3.40 Calcolo della probabilità di rispondere correttamente ad un <i>item</i>	169
Formula n. 3.41 Calcolo della probabilità di rispondere correttamente ad un <i>item</i> in logit	170
Formula n. 3.42 Calcolo dell' <i>Item Information Function</i>	173
Formula n. 3.43 Calcolo della <i>Test Information Function</i>	175
Formula n. 3.44 Calcolo del coefficiente di affidabilità	177
Formula n. 3.45 Calcolo della varianza totale	177
Formula n. 3.46 Calcolo del coefficiente di affidabilità 2	179
Formula n. 3.47 Calcolo del coefficiente di affidabilità di Guttman	179
Formula n. 3.48 Calcolo del coefficiente di affidabilità di Spearman-Brown	179
Formula n. 3.49 Calcolo del numero degli <i>item</i>	180

Formula n. 3.50 Calcolo del coefficiente <i>alpha</i>	180
Formula n. 3.51 Calcolo del coefficiente di affidabilità di Kuder-Richardson-20	180
Formula n. 3.52 Calcolo del coefficiente di affidabilità di Kuder-Richardson-21	181
Formula n. 3.53 Calcolo del coefficiente di affidabilità di Kuder-Richardson-21 2	181
Formula n. 3.54 Calcolo del coefficiente di affidabilità 3.....	182
Formula n. 3.55 Calcolo del coefficiente <i>alpha</i> 2	183
Formula n. 3.56 Calcolo dell'errore standard di misura (test normativi)	184
Formula n. 3.57 Calcolo dell'intervallo di confidenza.....	184
Formula n. 3.58 Calcolo della varianza totale (<i>crossed design</i>)	188
Formula n. 3.59 Calcolo della varianza totale (<i>nested design</i>)	188
Formula n. 3.60 Calcolo dell'indice di generalizzabilità	189
Formula n. 3.61 Calcolo dell'errore standard.....	189
Formula n. 3.62 Calcolo del coefficiente di affidabilità 4.....	190
Formula n. 3.63 Calcolo dell'errore standard di misura (test criteriali)	190
Formula n. 3.64 Calcolo dell'errore standard individuale.....	191
Formula n. 3.65 Calcolo della validità di contenuto	192
Formula n. 3.66 Calcolo della percentuale di risposte corrette (prove R/W)	194
Formula n. 3.67 Calcolo della percentuale di risposte corrette (prove P-C)	194
Formula n. 3.68 Calcolo del rango percentile	195

INDICE DELLE TABELLE

Tabella n. 1.1 - Livelli comuni di riferimento: scala globale	34
Tabella n. 1.2 - Scala di competenza.....	38
Tabella n. 1.3 - Equiparazione dei livelli ALTE e QCER	56
Tabella n. 1.4 - <i>Overall General Ability</i>	57
Tabella n. 1.5 - Esami Quadro ALTE per le certificazioni della lingua inglese	59
Tabella n. 3.1 - Esempi di distribuzione dei punteggi	127
Tabella n. 3.2 - Esempio di rappresentazione delle frequenze.....	143
Tabella n. 3.3 - Esempio di rappresentazione dei dati per il calcolo della deviazione standard.....	144
Tabella n. 3.4 - Esempi di distribuzione sulla base dell'indice di curtosi e del coefficiente di asimmetria.....	147
Tabella n. 3.5 - Relazioni tra variabili per potenza e direzione	150
Tabella n. 3.6 - Esempio di conteggio degli <i>item</i>	159
Tabella n. 3.7 - Esempio di calcolo della frequenza cumulativa	195
Tabella n. 3.8 - Esempio di calcolo del rango percentile	196
Tabella n. 4.1 - Composizione della prova di lettura PET	207
Tabella n. 4.2 - Composizione della prova di scrittura PET	207
Tabella n. 4.3 - Composizione della prova di ascolto PET	208
Tabella n. 4.4 - Composizione della prova di parlato PET.....	208
Tabella n. 4.5 - Composizione della prova di lettura PFLISP	209
Tabella n. 4.6 - Composizione della prova di ascolto PFLISP.....	209
Tabella n. 4.7 - Composizione della prova di scrittura PFLISP	210
Tabella n. 4.8 - Composizione della prova di parlato PFLISP	210
Tabella n. 4.9 - Rilevanza prove livello B1	211
Tabella n. 4.10 - Tempo concesso per le prove di livello B1	211
Tabella n. 4.11 - Indice di difficoltà degli <i>item</i> delle prime quattro prove di ascolto.....	216
Tabella n. 4.12 - Prova di ascolto n. 1 Parte II <i>Item</i> n. 1.....	216
Tabella n. 4.13 - Indice di difficoltà degli <i>item</i> della quinta prova di ascolto	217
Tabella n. 4.14 - Indice di difficoltà degli <i>item</i> delle prime quattro prove di lettura	217
Tabella n. 4.15 - Prova di lettura n. 1 Parte I <i>Item</i> n. 1 e n.2.....	218
Tabella n. 4.16 - Indice di difficoltà degli <i>item</i> della quinta prova di lettura	219
Tabella n. 4.17 - Prova di lettura n. 5 Parte II <i>Item</i> n. 1	219
Tabella n. 4.18 - Prova di ascolto n. 3 Parte II <i>Item</i> n. 1.....	220
Tabella n. 4.19 - Prova di ascolto n. 5 Parte II <i>Item</i> n. 3.....	221
Tabella n. 4.20 - Prova di lettura n. 1 Parte IV <i>Item</i> n. 2.....	221
Tabella n. 4.21 - Prova di lettura n. 5 Parte II <i>Item</i> n. 2	222
Tabella n. 4.22 - Indice di discriminazione degli <i>item</i> delle prime quattro prove di ascolto	222
Tabella n. 4.23 - Prova di ascolto n. 1 Parte II <i>Item</i> n. 5.....	223
Tabella n. 4.24 - Indice di discriminazione degli <i>item</i> della quinta prova di ascolto	224
Tabella n. 4.25 - Prova di ascolto n. 5 Parte I <i>Item</i> n. 2.....	224
Tabella n. 4.26 - Indice di discriminazione degli <i>item</i> delle prime quattro prove di lettura	224
Tabella n. 4.27 - Prova di lettura n. 4 Parte I <i>Item</i> n. 2	226
Tabella n. 4.28 - Indice di discriminazione degli <i>item</i> della quinta prova di lettura	226
Tabella n. 4.29 - Prova di lettura n. 5 Parte I <i>Item</i> n. 5	227
Tabella n. 4.30 - Percentuali degli indici presi in esame.....	232
Tabella n. 4.31 - Rappresentazione percentuale delle distribuzioni	235
Tabella n. 4.32 - Poligono delle frequenze delle cinque distribuzioni di punteggi	237
Tabella n. 4.33 - Coefficiente di correlazione di Spearman	238

ABSTRACT

La tesi si propone di discutere i processi di valutazione della competenza linguistico-comunicativa. La ricerca è stata condotta su una base teorica e sperimentale.

Nella prima parte il focus della riflessione riguarda l'esame delle principali teorie e metodologie che stanno alla base del testing linguistico. Nella seconda parte si è provveduto alla presentazione e all'analisi semplificata di alcune procedure statistiche idonee alla valutazione linguistica.

Scopo del lavoro è stato quello di scegliere i metodi di analisi quantitativa più adeguati per la valutazione della competenza linguistica e per l'adeguatezza delle prove, e di mostrarne un'applicazione pratica.

Il lavoro sperimentale ha preso come riferimento il progetto PFLISP, programma di valutazione delle competenze linguistiche acquisite da docenti della scuola primaria in vista dell'inserimento nelle liste di insegnanti di lingua inglese per la scuola primaria, svolto presso il Centro Linguistico di Ateneo dell'Università di Palermo. L'indagine statistica svolta sui dati ottenuti nelle prove somministrate durante un periodo di due mesi differenti (marzo e giugno) ha prodotto risultati diversi, che posti a confronto hanno mostrato margini di miglioramento.

In conclusione si è dimostrata la fondatezza degli strumenti di analisi e l'utilità che ne può derivare dall'utilizzo in ambito linguistico per un perfezionamento qualitativo dei test.

INTRODUZIONE

Le presente tesi si propone di illustrare l'importanza della conoscenza delle lingue straniere nella società di oggi, in particolare della lingua inglese; di spiegare i processi di valutazione della competenza linguistica dal punto di vista concettuale, attraverso alcune teorie sul *testing* linguistico, e dal punto di vista empirico, tramite l'analisi statistica di dati ricavabili da prove linguistiche.

L'interesse per la valutazione linguistica, durante gli anni del dottorato di ricerca, ha condotto alla frequenza di corsi sulla valutazione e sulla docimologia nel territorio italiano, favorendo l'approfondimento degli studi sul *testing* linguistico mediante la partecipazione a corsi organizzati dall'ALTE (*Association of Language Testers in Europe*); la ricerca di contatti con esperti di valutazione a livello europeo per indicazioni bibliografiche e consulenze specialistiche sulle teorie e pratiche attualmente in vigore; lo studio delle stesse e la partecipazione a progetti attinenti alla preparazione e allo sviluppo di prove linguistiche e all'analisi dei risultati conseguiti.

Il percorso di ricerca parte dallo studio delle teorie sulla valutazione linguistica per giungere all'esame delle applicazioni pratiche di metodi teorici e statistici, condotte su prove empiriche. Il lavoro è diviso in quattro parti.

Il primo capitolo, *Educazione plurilinguistica e certificazione*, mette in evidenza il ruolo chiave svolto dalle lingue europee nell'ambito dei legami linguistici, storici e culturali esistenti tra i paesi dell'Unione Europea e tra l'UE e paesi terzi.

Viene presentata una documentata analisi delle politiche educative

affermatesi a livello europeo e internazionale a sostegno della promozione dell'apprendimento delle lingue e dell'educazione linguistica, che mette in luce come gli orientamenti assunti dagli organismi comunitari abbiano influito sulle decisioni dei diversi Paesi in materia di educazione e ricerca. Si evince così un rapporto intenso tra politica e ricerca. Da una parte, attraverso documenti e iniziative, si preme sull'educazione plurilinguistica e dall'altra parte ci si interroga sulle trasformazioni che subiscono i concetti di "lingua", "competenza", "performance" e "valutazione". Si cercano dunque nuove metodologie quantitative e qualitative, che fungano da sostegno per lo sviluppo di pratiche efficaci per la valutazione delle competenze linguistiche, intese come conoscenze di quelle "pratiche linguistiche" che permettono a un individuo di interagire in un contesto situazionale in cui la lingua veicolo è diversa dalla lingua materna.

Questo approccio è iniziato negli anni Settanta e Ottanta con la pubblicazione, da parte del Consiglio d'Europa, del "livello soglia", una lista di "funzioni comunicative" riferite alle varie lingue principali (ad es. il livello soglia risale al 1955 per l'inglese e al 1982 per l'italiano), elaborata con l'obiettivo di fornire ai cittadini europei una padronanza linguistica tale da consentire di sopravvivere in un paese straniero.

Tale orientamento è proseguito con la pubblicazione, nel 2001, del *Quadro Comune Europeo di Riferimento per le Lingue*, in cui si stabiliscono i livelli comuni di riferimento. La presenza di tali livelli di riferimento, necessaria per sostenere e restare al passo con le politiche educative europee, ha portato allo sviluppo di un maggiore interesse, da parte di enti certificatori e studiosi del *testing* linguistico, verso gli strumenti valutativi e le pratiche di misurazione. L'attenzione a questi ultimi è dovuta dal desiderio sempre più crescente di voler

dimostrare, da parte degli individui, la propria competenza a livello europeo e sfruttare la spendibilità di una certificazione riconosciuta e standardizzata in ambito scolastico e/o professionale.

Se il primo capitolo esplicita la ragione per un interesse attivo verso la valutazione e la certificazione linguistica, con il secondo capitolo, *Testing linguistico e valutazione*, si entra nel merito di alcune delle teorie che sottostanno allo sviluppo di metodologie per la progettazione di prove linguistiche e per l'analisi dei risultati ottenuti da tali prove.

Il modello di Weir è la dimostrazione della grande trasformazione, in campo educativo e valutativo, del modo di concepire la conoscenza e l'uso della lingua, e delle implicazioni sociali che ne derivano.

Le prove linguistiche giocano oggi un importante ruolo nella vita degli individui, sia nel campo dell'educazione che della professione e della mobilità da un paese all'altro; di conseguenza occorrono prove empiriche a dimostrazione della validità e dell'affidabilità dei risultati ottenuti e certificati relativamente alle competenze linguistiche possedute o raggiunte da un individuo.

Nel terzo capitolo, *Statistica e analisi dei dati*, vengono introdotti e analizzati alcuni strumenti di analisi desunti dalla statistica descrittiva. Le procedure illustrate sono funzionali alla misurazione dei risultati, la quale si rende necessaria per due ragioni: per una verifica dell'appropriatezza del contenuto delle prove rispetto allo scopo per cui le stesse sono state progettate, e per una valutazione efficace della competenza linguistica del candidato.

I metodi e le formule statistiche presentate nel capitolo, applicabili per la verifica dell'efficacia degli *item* sulla misurazione delle *performance* dei candidati per un determinato livello di competenza linguistica, fanno riferimento a due

modelli teorici, la *Classical Test Theory* e la *Item Response Theory*.

Le procedure statistiche, illustrate per l'interpretazione delle *performance* dei soggetti sottoposti a esame, conducono al calcolo di misure della tendenza centrale e della dispersione o variabilità. Si può affermare difatti che un punteggio di per sé non ha alcun significato finché non viene posto a confronto con altri punteggi.

Nel quarto capitolo, *Un caso di studio: PFLISP Sicilia*, si entra nel cuore dell'analisi, poiché si prendono in esame dati reali. Innanzitutto è importante sottolineare che un caso di studio è l'esame di uno specifico fenomeno, un esempio di una qualche ipotesi, problema o realtà. È uno studio che consente dapprima di generare ipotesi e poi, dopo averle testate, di confutarle o avvalorarle. Piuttosto che analizzare un numero limitato di variabili mediante l'esame di un campione di individui per inferire sull'intera popolazione, un caso di studio è un'analisi approfondita di un singolo caso concreto inserito in un particolare contesto. Tale metodo consente di raccogliere i dati, analizzarli e presentarne i risultati in modo empirico e sistematico, ossia di effettuare una vera e propria analisi sul campo. Esso consente di ottenere la struttura statistica per avanzare inferenze partendo da dati quantitativi reali.

I dati concreti sui quali verte il caso di studio del presente lavoro provengono dalle prove concorsuali somministrate presso il CLA dell'Università di Palermo all'interno del "Piano di Formazione per lo Sviluppo delle Competenze linguistico-comunicative e metodologico-didattiche in Lingua Inglese degli insegnanti di Scuola Primaria" (PFLISP) avviato dal Ministero dell'Istruzione, dell'Università e della Ricerca.

Il progetto PFLISP Sicilia, coordinato dal CLA di Palermo, ha interessato anche i CLA di Catania, Enna e Messina. Il progetto, che ha portato alla

formulazione di criteri per lo svolgimento dei corsi e delle prove, all'elaborazione dei materiali per le prove e alla correzione delle stesse, è stato seguito da chi scrive in tutte le fasi, soprattutto per quanto riguarda la registrazione dei brani e dei dialoghi, il trasferimento del materiale sulla piattaforma *e-learning* Anacleto e l'analisi statistica dei risultati. Lo scopo di tale analisi consiste nel dar prova dell'utilità dei metodi statistici attraverso l'applicazione di alcune procedure analitiche (esposte nel terzo capitolo) su un caso di sperimentazione concreta.

Nelle conclusioni vengono riepilogati i risultati dell'analisi dei dati provenienti dalle prove PFLISP e le considerazioni emerse a seguito della rilevazione, e si suggeriscono alcune proposte di ricerca legate all'argomento.

Infine in appendice sono illustrate alcune tabelle desunte dal QCER in originale (appendice A), un *sample test* PFLISP (appendice B) e i dati (calcolati mediante le necessarie formule statistiche) ricavati dalle cinque prove PFLISP somministrate durante le sessioni d'esame presso il CLA di Palermo nell'anno 2009 (appendice C).



Allo scopo di favorire la lettura di questo lavoro si indicano qui di seguito le convenzioni utilizzate:

- le doppie virgolette “abcd” indicano, in genere, una citazione o un’etichetta;
- le parentesi quadre [abcd] all’interno delle citazioni indicano le parole o espressioni non contenute nell’originale;
- i tre punti di sospensione tra parentesi quadre [...] indicano l’omissione di una parola o una parte di frase di una citazione presente nell’originale;
- il carattere corsivo *abcd* (escludendone l’ovvio utilizzo per i titoli delle opere) indica termini stranieri, espressi nella lingua d’origine;

- il carattere maiuscolo (escludendone il prevedibile utilizzo nelle iniziali e a inizio di frase) indica una sigla o un'abbreviazione (le abbreviazioni più comuni sono L2 per indicare una lingua seconda, QCER per indicare il *Quadro Comune Europeo di Riferimento per le Lingue*, UE per l'Unione Europea e CE per Comunità Europea);
- le parentesi tonde (escludendone il prevedibile utilizzo per racchiudere una spiegazione o un commento) indicano i riferimenti bibliografici o i rimandi a figure, tabelle, grafici e formule inseriti nella tesi;
- le note a piè di pagina indicano, in linea di massima, riferimenti per approfondimenti sull'argomento trattato.

Altresì si indicano di seguito alcune precisazioni di carattere lessicale sulle scelte compiute laddove alcuni concetti chiave risultano problematici per la glottodidattica:

- il termine “abilità” nel QCER è connesso con le competenze generali. Per la glottodidattica italiana le “abilità” sono il saper ascoltare, parlare, leggere e scrivere (dette abilità primarie) e le abilità integrate quali il saper interpretare, parafrasare, tradurre¹. Il QCER si riferisce alle abilità nel senso glottodidattico attraverso il termine “attività”, il quale, in glottodidattica, si riferisce invece a un significato più polisemico che abbraccia i compiti svolti dall'apprendente, le tecniche didattiche utilizzate dal docente, ecc. Qui, per evitare ogni incomprensione, si preferisce utilizzare il termine “abilità” per riferirsi alle abilità primarie e integrate, anche quando il QCER fa riferimento al termine “attività”. Ci si riferisce invece ad “attività” per designare diversi tipi di abilità, anche non linguistiche;
- i termini “competenza” e “conoscenza” linguistica sono utilizzati qui in modo intercambiabile e con essi ci si riferisce al concetto di competenza comunicativa così come inteso da Hymes, ossia la capacità di servirsi di tutti i codici, verbali e non, per svolgere un atto comunicativo (“sapere la lingua” e “saper fare con la lingua”), e al

¹ Cfr. Balboni, 1999, s.v. “abilità”.

concetto di “saper fare lingua”, ovvero alla padronanza dei processi cognitivi, oltre che linguistici, che sottostanno alle abilità linguistiche².

Si precisa inoltre che le leggi e i documenti emanati da organismi europei a cui si fa riferimento in questo lavoro sono citati in lingua italiana se prodotti o pubblicati in tale lingua. Per i documenti non disponibili in lingua italiana, ci si riferisce alla versione redatta in lingua inglese.

Infine l'esistenza e la funzionalità degli indirizzi web indicati nei vari capitoli dell'intero lavoro sono verificate fino al giorno 10 del mese di febbraio dell'anno 2011.

² Per una definizione più dettagliata di ciò che si intende per competenza comunicativa, “sapere la lingua”, “saper fare con la lingua”, “saper fare lingua” e “saper integrare la lingua con i linguaggi non verbali” si rimanda al paragrafo 1.3.1 del presente lavoro e a Balboni, 2002:73.

Educazione plurilinguistica e certificazione

1.1 Premesse introduttive

L'educazione linguistica è un processo che si realizza attraverso l'insegnamento e l'apprendimento delle lingue, dalla lingua materna alle lingue seconde. L'espressione "lingua seconda" (L2) si riferisce al fatto che la lingua in questione viene appresa in un momento successivo rispetto a quello in cui è stata appresa la lingua madre o prima lingua (L1). Spesso però con L2 ci si riferisce a una lingua che non è cronologicamente la seconda, ma la terza o quarta lingua e così via (Rod, 1997:3). L'espressione "apprendimento delle lingue" si riferisce al processo attraverso cui gli individui imparano una varietà di lingue in aggiunta alla loro lingua nativa, all'interno o all'esterno di un contesto scolastico; e con l'espressione SLA (*Second Language Acquisition*) si indica lo studio di tale processo. Il termine "acquisizione" è stato introdotto da Krashen in un'ipotesi dicotomica tra acquisizione e apprendimento¹. Secondo lo studioso l'acquisizione è un processo automatico e spontaneo (stabile) che afferisce alla L1 e alle L2 apprese in contesti naturali, e l'apprendimento è un processo razionale e intenzionale (di durata breve) che afferisce alle lingue straniere (LS) apprese in contesti formali. Tale ipotesi è stata spesso contestata poiché è quasi impossibile che ci sia una netta distinzione tra i due processi e che l'apprendimento non possa talvolta trasformarsi in acquisizione. Pertanto alcuni

¹ Qui i termini "apprendimento" e "acquisizione" verranno utilizzati interscambiabilmente.

autori propongono il termine iperonimo “appropriazione” per superare questa separazione (Bosisio, 2005a:97-98). Una seconda distinzione da fare è relativa ai concetti L2 e LS: per L2 si intende una lingua che viene studiata in un’area geografica in cui essa è parlata (dunque in contesti non necessariamente scolastici), e con LS si intende una lingua che viene studiata in un’area geografica in cui essa non è presente (dunque esclusivamente in contesti scolastici)².

L’apprendimento di una lingua altra può avere dunque luogo a scuola ma può interessare altri aspetti della vita di un individuo.

“What comes to mind for many people when they encounter the phrase “second language acquisition”, is the experience they had as school students when they were engaged in the study of one or more foreign languages. [...].

The acquisition of second languages in a formal school setting, however, is not the only context where second languages have their place in the world today. [...].

In the interdependent world of today, second language acquisition and use are ubiquitous.” (Larsen-Freeman e Long, 1991:1-2)

Per venire incontro al crescente bisogno di comunicazione internazionale, l’apprendimento di una L2 oggi sembra essere una questione di primaria importanza al di là di qualunque contesto scolastico.

Secondo Ellis Rod un vero e proprio studio sistematico del processo di apprendimento di una lingua seconda si è sviluppato intorno alla seconda metà del ventesimo secolo congiuntamente alla nascita del World Wide Web e del

² Qui non verrà fatta tale distinzione e, da questo momento, verranno utilizzati interscambiabilmente i termini lingua seconda, lingua straniera e lingua altra per indicare una lingua diversa da quella materna. L’espressione “lingua altra” va intesa come iperonimo di lingua straniera e di lingua seconda (Bosisio 2007a:377). Per maggiori approfondimenti si rimanda a Balboni, 2002:58-60.

villaggio globale³, ossia al sorgere dell'esigenza di comunicazione tra individui appartenenti a comunità diverse.

Sostanzialmente si potrebbe affermare che negli ultimi anni vi è stato un cambiamento di prospettiva nell'ambito dell'apprendimento linguistico. Il focus si è spostato, dal punto di vista dell'apprendimento, sul processo comunicativo della lingua (piuttosto che sulla lingua come sistema linguistico), e dal punto di vista dell'insegnamento, sul soggetto apprendente (piuttosto che sull'oggetto da apprendere). Questo cambiamento di prospettiva ha dato l'avvio a studi e ricerche nel campo della SLA: l'apprendimento linguistico fuoriesce dal campo ristretto della classe o della scuola, per propagarsi all'esterno, interessando svariati versanti della vita di un soggetto. Il contesto scolastico resta il luogo principale per lo sviluppo di un momento di insegnamento/apprendimento linguistico, in cui l'insegnante e l'apprendente rimangono gli attori principali. Il modello didattico tradizionale propone una rappresentazione tripartita in cui figurano l'insegnante, l'apprendente e l'oggetto dell'apprendimento rispettivamente ai tre vertici di un triangolo (fig. n. 1.1).

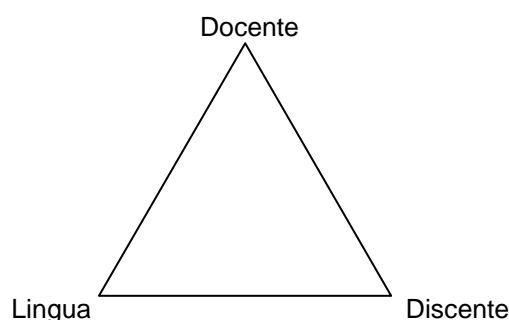


Figura n. 1.1 – Il tradizionale modello di atto didattico (Larsen-Freeman e Long, 1991:3)

L'apprendente rappresenta il discente, italofono e non, che fa esperienza

³ Con l'espressione "villaggio globale" Marshall McLuhan nel 1964 intendeva un mondo in cui, grazie alle tecnologie e alla loro influenza sullo stile di vita degli individui, si annullano le distanze fisiche e culturali.

di una L2 nella scuola, sia come materia prevista dai programmi scolastici, sia come strumento per comunicare in classi plurilingue (sempre più numerose⁴) con i propri compagni.

L'educazione linguistica oggi però ha luogo in una realtà multiforme, delineata in modo esemplare dal "modello della costellazione didattica" di Dabène (fig. n. 1.2).



Figura n. 1.2 – La costellazione didattica

Con la metafora della costellazione, Dabène evidenzia le numerose variabili che entrano in gioco in un atto didattico. Oltre agli elementi già presenti nel tradizionale modello del triangolo didattico, l'oggetto d'insegnamento/apprendimento (lingua), il soggetto (discente) e l'agente (docente), appare un quarto elemento in cui si esplica l'apprendimento: il contesto educativo e sociale (Bosisio, 2007a:380; Bosisio, 2005b:220).

L'educazione linguistica, rivolgendosi a un pubblico eterogeneo, dovrebbe

⁴ Secondo il XIX rapporto annuale sull'immigrazione di Caritas/Migrantes del 2009 sono circa 6 mila gli studenti stranieri che si laureano annualmente in Italia (<http://www.dossierimmigrazione.it/schede/pres2009--pittau.pdf>, ultima consultazione: febbraio 2011) e nell'anno scolastico 2008/2009 gli alunni figli di genitori stranieri erano 628.937 su un totale di 8.943.796 iscritti. L'aumento annuale è stato di 54.800 unità, pari a circa il 10% (<http://www.dossierimmigrazione.it/schede/pres2009--scheda.pdf>, ultima consultazione: febbraio 2011).

tendere allo sviluppo e alla valorizzazione del plurilinguismo, al cui tema appare sensibile l'Unione Europea.

1.2 Una realtà europea per le lingue

L'Unione Europea di fatto è stata, ed è tuttora, promotrice di diverse iniziative volte a incoraggiare l'istruzione e la formazione, soprattutto linguistica, dei cittadini europei.

Insieme al Consiglio d'Europa, l'Unione Europea si è fatta fautrice, nel 2001⁵, dell'**anno europeo delle Lingue**, il cui scopo è stato quello di incoraggiare lo studio delle lingue straniere e promuovere la diversità linguistica europea.

La politica dell'Unione tende da una parte a promuovere la diversità tra i paesi e dall'altra parte a ridurla. Sembrerebbe una contraddizione, in realtà le attività dell'Unione Europea sono volte a garantire la pace, la prosperità e la stabilità dei popoli, favorendo uno sviluppo economico e sociale equilibrato, che tende a ridurre il divario fra ricchi e poveri; non dimenticando di preservare però le differenze culturali e linguistiche dei paesi, di fronte a un mondo che volge alla globalizzazione.

La **Carta dei diritti fondamentali**⁶ dell'Unione Europea, proclamata nel 2000, che sancisce tutti i diritti attualmente riconosciuti dagli Stati membri e dai loro cittadini, garantisce, secondo l'articolo 22 (*"L'Unione rispetta la diversità culturale, religiosa e linguistica"*), il rispetto della diversità linguistica, che non

⁵ Nello stesso anno è stata istituita la **giornata europea delle lingue**, che si celebra il 26 settembre.

⁶ La *Carta dei diritti fondamentali* è disponibile online all'indirizzo http://www.europarl.europa.eu/charter/pdf/text_it.pdf (ultima consultazione: febbraio 2011).

dovrebbe compromettere, dunque, le specificità regionali e la ricchezza delle diverse tradizioni e culture.

La tutela della diversità linguistica, eredità culturale del nostro patrimonio comune da salvaguardare, è prerogativa dell'Unione Europea. Di diversa natura è la questione della comunicazione linguistica nelle istituzioni comunitarie e tra i cittadini europei, di competenza della Comunità Europea, che si interessa di garantire la mobilità professionale e personale all'interno dell'Europa. L'articolo 149 del **Trattato che istituisce la Comunità europea**⁷ afferma che *“La Comunità contribuisce allo sviluppo di un'istruzione di qualità incentivando la cooperazione tra Stati membri [...]. L'azione della Comunità è intesa: a sviluppare la dimensione europea dell'istruzione, segnatamente con l'apprendimento e la diffusione delle lingue degli Stati membri; a favorire la mobilità degli studenti e degli insegnanti [...]. La Comunità e gli Stati membri favoriscono la cooperazione con i paesi terzi e le organizzazioni internazionali competenti in materia di istruzione, in particolare con il Consiglio d'Europa.”*

1.2.1 Panoramica storica

La Comunità Europea, il Consiglio d'Europa e altre numerose istituzioni, comunitarie e non, hanno mostrato particolare interesse per la questione linguistica europea, impegnandosi in tal senso nello sviluppo di iniziative e documenti significativi.

Il Consiglio d'Europa è stato senza dubbio il primo organismo a dare un forte impulso alla promozione dello studio delle lingue, della storia e della civiltà

⁷ Il *Trattato che istituisce la Comunità europea* è disponibile online all'indirizzo http://eur-lex.europa.eu/it/treaties/dat/12002E/pdf/12002E_IT.pdf (ultima consultazione: febbraio 2011).

dei paesi, alla salvaguardia della cultura europea e delle diversità nazionali e allo sviluppo di una reciproca comprensione tra i popoli europei, attraverso la **Convenzione culturale europea**⁸ del 1954.

Il secondo documento degno di nota sembra essere il progetto **Lingue Moderne/Lingue Vive**⁹, approvato dal Consiglio d'Europa nel 1975, che prevedeva la pubblicazione del **Threshold Level**, un inventario di nozioni e risorse utili per il raggiungimento del "Livello Soglia", ossia una competenza comunicativa di base. Il raggiungimento del livello soglia attestava una padronanza tale da consentire a un adulto di interagire in contesti di vita quotidiana in un paese straniero.

Lavori rilevanti che hanno contribuito al raggiungimento di decisioni importanti in ambito linguistico sono state una serie di **Raccomandazioni e Risoluzioni**¹⁰ del Consiglio d'Europa. Di particolare importanza sono le *Raccomandazioni dei Comitati dei Ministri* del 1982 (n. 18) e del 1998 (n. 6), le quali precisano come nell'apprendimento linguistico i concetti di plurilinguismo e pluriculturalismo assumano sempre più rilevanza. Tali Raccomandazioni sono state fondamentali per la realizzazione del **QCER - Quadro comune europeo di riferimento per le lingue** (sul cui contenuto si approfondirà in seguito) pubblicato nel 2001.

Vi sono state, inoltre, grazie al lavoro della Commissione Europea, iniziative notevoli, dal 1990 al 1994 tramite il **Programma Lingua**, rafforzato in seguito da programmi analoghi, quali ad esempio Comenius, Erasmus, Leonardo

⁸ La *Convenzione culturale europea* è consultabile online all'indirizzo <http://www.admin.ch/ch/i/rs/i4/0.440.1.it.pdf> (ultima consultazione: febbraio 2011).

⁹ La doppia denominazione deriva dal titolo del progetto inglese (*Modern Language*) e da quello francese (*Langues Vivantes*).

¹⁰ Per una panoramica delle più importanti Raccomandazioni e Risoluzioni si rinvia a Langé, 2005:84-85. La lista completa delle Raccomandazioni e Risoluzioni è consultabile online all'indirizzo http://www.coe.int/t/i/cplre/testi_in_italiano/ (ultima consultazione: febbraio 2011).

da Vinci e Socrates solo per citarne alcuni, che contribuiscono, attraverso la mobilità di studenti e insegnanti all'interno dei paesi membri, alla promozione della conoscenza delle lingue straniere in ambito europeo.

La Commissione Europea ha apportato il proprio contributo in questo ambito, oltre che con progetti d'azione come quelli sopra accennati, attraverso numerosi documenti di riflessione, i quali introducono cambiamenti nella vita di coloro che apprendono e di coloro che insegnano le lingue. Esempi di alcune iniziative recenti sono il **Libro bianco - Istruzione e formazione - Insegnare e apprendere - Verso la società conoscitiva**¹¹, del 1995, in cui, per la prima volta, la Commissione identifica, tra i cinque obiettivi generali, il conseguimento da parte di tutti i cittadini dell'UE della padronanza di tre lingue europee, due lingue comunitarie in aggiunta alla lingua materna secondo la formula 1+>2 (Mackiewicz, 2004:176). I Libri bianchi sono documenti ufficiali, pubblicati annualmente dalla Commissione Europea, che contengono proposte di azione comunitaria in vari settori. Ogni anno la Commissione redige, in aggiunta, i Libri verdi, tramite i quali si illustra lo stato di un determinato settore da disciplinare. Nel 1996, attraverso il **Libro verde - Istruzione su Formazione e Ricerca - Gli ostacoli alla mobilità transnazionale**¹², e successivamente, nel 2008 con **Migrazione e mobilità: le sfide e le opportunità per i sistemi d'istruzione europei**¹³ e nel 2009 con **Promuovere la mobilità dei giovani per**

¹¹ Il *Libro bianco - Istruzione e formazione - Insegnare e apprendere - Verso la società conoscitiva* è consultabile online in lingua inglese all'indirizzo http://europa.eu/documents/comm/white_papers/pdf/com95_590_en.pdf (ultima consultazione: febbraio 2011).

¹² Il *Libro verde - Istruzione su Formazione e Ricerca - Gli ostacoli alla mobilità transnazionale* è consultabile online in lingua inglese all'indirizzo http://aei.pitt.edu/1226/01/education_mobility_obstacles_gp_COM_96_462.pdf (ultima consultazione: febbraio 2011).

¹³ Il *Libro verde - Migrazione e mobilità: le sfide e le opportunità per i sistemi d'istruzione europei* è consultabile online all'indirizzo <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2008:0423:FIN:IT:PDF> (ultima consultazione: febbraio 2011).

*l'apprendimento*¹⁴, la Commissione si interessa di facilitare gli scambi tra gli Stati membri, nella convinzione che “*La mobilità per l'apprendimento, ovvero la mobilità transnazionale volta all'acquisizione di nuove competenze, è uno dei mezzi fondamentali attraverso i quali una persona, soprattutto se giovane, può incrementare le proprie possibilità di occupazione e potenziare il proprio sviluppo personale. [...] La mobilità va collegata a specifici risultati dell'apprendimento per l'ottenimento di qualifiche, crediti e/o esperienza professionale. [...] Alcune delle principali competenze da acquisire in un periodo di mobilità a fini di apprendimento sono le lingue straniere e le competenze interculturali.*” (Promuovere la mobilità dei giovani per l'apprendimento, 2009:2-4).

Un ulteriore passo importante si è compiuto nel 1997 con la Conferenza di Lussemburgo, **Early Learning and After**, organizzata dalla Commissione Europea, durante la quale, oltre a ribadire l'estrema rilevanza della competenza plurilingue, è stato vivamente sostenuto l'apprendimento precoce delle lingue già nelle scuole materne ed elementari.

L'anno successivo la Commissione Europea si è fatta promotrice del progetto **Label Europeo**, un riconoscimento che viene assegnato annualmente alle proposte più innovative realizzate nell'ambito delle lingue straniere. L'iniziativa LABEL tende a incoraggiare il “*Lifelong Learning*”¹⁵ (apprendimento permanente), a incoraggiare la realizzazione di progetti linguistici innovativi e a sensibilizzare gli individui alla diversità linguistica e culturale.

Nel 2000 il Consiglio Europeo, a Lisbona, traccia le linee per una “*Unione [che] si è prefissata un nuovo obiettivo strategico per il nuovo decennio:*

¹⁴ Il *Libro verde - Promuovere la mobilità dei giovani per l'apprendimento* è consultabile online all'indirizzo <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0329:FIN:IT:PDF> (ultima consultazione: febbraio 2011).

¹⁵ Il Lifelong Learning Programme (LLP) è stato istituito dal Parlamento europeo e dal Consiglio nel 2006.

diventare l'economia basata sulla conoscenza più competitiva e dinamica del mondo, in grado di realizzare una crescita economica sostenibile con nuovi e migliori posti di lavoro e una maggiore coesione sociale.” (Conclusioni della Presidenza, Consiglio Europeo, Lisbona, 2000). Ci si prefigge inoltre di diminuire il tasso di abbandono scolastico per combattere il fenomeno della dispersione scolastica; di definire un quadro europeo in cui le nuove competenze di base da fornire lungo tutto l'arco della vita dovrebbero comprendere “competenze in materia di tecnologie dell'informazione, lingue straniere, cultura tecnologica, imprenditorialità e competenze sociali; [...] [di] promuovere la mobilità di studenti, docenti e personale preposto alla formazione e alla ricerca, sia utilizzando al meglio i programmi comunitari esistenti (Socrates, Leonardo, Gioventù) eliminando gli ostacoli, sia mediante una maggiore trasparenza nel riconoscimento delle qualifiche e dei periodi di studio e formazione [...].” (Conclusioni della Presidenza, Consiglio Europeo, Lisbona, 2000).

Sulla scia degli obiettivi delineati nel Consiglio di Lisbona, in un processo di europeizzazione dei metodi di insegnamento e di apprendimento delle lingue e culture d'Europa e di certificazioni europee, nel 2001 (in occasione dell'Anno Europeo delle Lingue), il Consiglio d'Europa appronta il sopraccitato **Quadro comune europeo di riferimento per le lingue**¹⁶ (il documento è stato pubblicato nelle versioni inglese e francese per la prima volta nel 1996, in italiano nel 2002). Si tratta del documento che conclude il lavoro di un gruppo di ricerca linguistica e pedagogica, attivato dal Consiglio d'Europa nel 1971, e apre una produttiva collaborazione fra gli stati comunitari in materia di apprendimento/insegnamento delle lingue moderne. L'importanza di questo documento è data dal fatto che

¹⁶ Il *Quadro comune europeo di riferimento per le lingue* è reperibile online in lingua inglese all'indirizzo http://www.coe.int/t/dg4/linguistic/source/Framework_EN.pdf (ultima consultazione: febbraio 2011).

costituisce una base comune (tanto che non esistono versioni diverse nelle varie lingue ma traduzioni dello stesso testo) per l'elaborazione di materiali, programmi, esami e certificazioni nell'ambito dell'educazione linguistica. Nell'ottica della nuova politica linguistica dell'UE (volta a promuovere il plurilinguismo), i diversi sistemi di certificazione devono, infatti, poter essere coordinati per mezzo di un quadro comune approvato in tutta Europa e oltre, al fin di poter garantire mutua comparabilità e agevolazioni per la mobilità (nel seguito del capitolo il documento in questione sarà analizzato in modo più esauriente).

Contestualmente, connesso al QCER, è stato presentato dal Consiglio d'Europa, il **Portfolio Europeo delle Lingue** (P.E.L.), un documento che accompagna coloro che studiano una lingua nel loro percorso di apprendimento per tutta la durata della vita. È uno strumento personale per l'apprendimento linguistico che permette la registrazione del percorso di apprendimento di uno studente, di porre nuovi obiettivi definendo e programmando le tappe del proprio apprendimento (il P.E.L. sarà approfondito opportunamente in seguito).

Nel 2003, considerando le richieste del Parlamento europeo, la Commissione europea ha presentato il **Piano d'azione 2004-2006**, destinato a promuovere la diversità linguistica e l'apprendimento delle lingue. *“L'obiettivo chiave [...] del documento consiste nell'estendere a tutti i cittadini i vantaggi dell'apprendimento delle lingue. Le conoscenze linguistiche sono infatti parte integrante delle competenze fondamentali che ogni cittadino deve avere ai fini della formazione, dell'occupazione, degli scambi culturali e della realizzazione personale. L'apprendimento delle lingue è un'attività che dura tutta la vita.”*¹⁷ La

¹⁷ Il *Promuovere l'apprendimento delle lingue e la diversità linguistica: Piano d'azione 2004-2006* è reperibile online all'indirizzo <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2003:0449:FIN:IT:PDF> (ultima consultazione: febbraio 2011).

Commissione europea garantisce, mediante siffatto Piano d'azione, che la promozione dell'apprendimento delle lingue e della diversità linguistica conservi la stessa importanza, preservata già dai programmi Socrates e Leonardo da Vinci negli anni precedenti, mediante i programmi successivi. Incoraggia, inoltre, gli Stati membri "a sfruttare le opportunità offerte da altri programmi europei, dal Fondo sociale europeo e dalla Banca europea per gli investimenti per finanziare nuove iniziative nel campo dell'apprendimento delle lingue." (*Promuovere l'apprendimento delle lingue e la diversità linguistica: Piano d'azione 2004-2006*).

In seguito al Piano d'azione 2004-2006, la Commissione europea, nel 2005, chiarisce ulteriormente il proprio impegno in favore della promozione del multilinguismo attraverso i documenti **L'indicatore europeo di competenza linguistica**¹⁸ e **Un nuovo quadro strategico per il multilinguismo**¹⁹. "Il multilinguismo si riferisce sia alla capacità del singolo di usare più lingue sia alla coesistenza di differenti comunità linguistiche in una determinata area geografica. Nel presente documento il termine viene impiegato per descrivere il nuovo settore d'intervento comunitario inteso a promuovere un clima favorevole alla piena espressione di tutte le lingue creando condizioni ottimali per l'insegnamento e l'apprendimento di diverse lingue." (*Un nuovo quadro strategico per il multilinguismo*).

Per confermare la valorizzazione della diversità linguistica, l'importanza del multilinguismo e dell'educazione linguistica, il Parlamento europeo si è fatto promotore, nel marzo del 2009, del documento **Multilinguismo: una risorsa per**

¹⁸ *L'indicatore europeo di competenza linguistica* è disponibile online all'indirizzo <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0356:FIN:IT:PDF> (ultima consultazione: febbraio 2011).

¹⁹ *Un nuovo quadro strategico per il multilinguismo* è reperibile online all'indirizzo <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0596:FIN:IT:PDF> (ultima consultazione: febbraio 2011).

l'Europa e un impegno comune²⁰. Con tale atto il Parlamento ha ribadito la necessità di riconoscere la parità tra le lingue ufficiali dell'Unione europea; ha riconfermato l'importanza dell'apprendimento di altre lingue dell'Unione europea, una delle quali dovrebbe essere la lingua di un paese vicino e l'altra una "lingua franca" internazionale, che possa fornire ai cittadini competenze e qualifiche per poter partecipare alla società democratica, in termini di cittadinanza attiva, occupabilità e conoscenza di altre culture; ha incoraggiato la salvaguardia delle lingue regionali e minoritarie (patrimonio culturale da proteggere e coltivare) e del multilinguismo nei paesi o nelle regioni in cui convivono più lingue ufficiali; e soprattutto ha richiamato l'attenzione sul divario comunicativo tra individui con culture diverse e tra persone che parlano più lingue e persone che ne parlano una sola, sottolineando il fatto che la carenza di competenze linguistiche costituisce ancora un grave ostacolo all'integrazione sociale e occupazionale. In tale prospettiva il Parlamento, in accordo con la Commissione, ha sottolineato la promozione della "lingua materna più altre due lingue" in ambito scolastico; incoraggiando l'apprendimento di una seconda lingua sia in età precoce sia in età adulta.

Questo documento riassume e ribadisce i passi compiuti sul piano dell'istruzione, dell'educazione linguistica, del multilinguismo, dell'integrazione e della coesione sociale, in cinquanta anni di impegno e collaborazione da parte delle istituzioni comunitarie e degli Stati membri. Inoltre sono incluse richieste per un ulteriore miglioramento futuro nel campo della conoscenza delle lingue e per l'istituzione di sistemi di convalida delle competenze linguistiche, sulla base del Quadro comune europeo di riferimento per le lingue.

²⁰ Il *Multilinguismo: una risorsa per l'Europa e un impegno comune* è reperibile online all'indirizzo <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P6-TA-2009-0162+0+DOC+XML+V0//it> (ultima consultazione: febbraio 2011).

Tale invito è ulteriormente ribadito nel documento del maggio 2009, un **Quadro strategico per la cooperazione europea nel settore dell'istruzione e della formazione**²¹, attraverso il quale il Consiglio “*conviene che nel periodo fino al 2020, l'obiettivo principale della cooperazione europea dovrebbe essere quello di sostenere l'ulteriore sviluppo dei sistemi di istruzione e formazione degli Stati membri*” (*Quadro strategico per la cooperazione europea nel settore dell'istruzione e della formazione*).

1.3 Il Quadro comune europeo di riferimento per le lingue

Il *Quadro comune europeo di riferimento per le lingue* nasce dall'esigenza di fornire uno strumento teorico, e al tempo stesso pratico, per la determinazione di *standard* di misurazione stabili e condivisi, equiparabili a livello internazionale.

È frutto di un lavoro iniziato negli anni Settanta, che ha visto la partecipazione di numerosi individui e istituzioni, coinvolti in un processo di ricerca scientifica, di consultazione e cooperazione, e rispecchia gli obiettivi ai quali mirano tutti i lavori e le iniziative comunitarie compiute nell'arco di poco più di cinquanta anni.

Esso provvede a una convergenza, favorendo la cooperazione tra le istituzioni educative europee, delle politiche nazionali degli Stati membri nel campo dell'apprendimento e dell'insegnamento delle lingue a livello europeo. Provvede soprattutto a fornire una base comune per la comparazione delle certificazioni linguistiche e per il riconoscimento reciproco dei titoli di studio, agevolando così la mobilità sul piano educativo e professionale.

²¹ Il *Quadro strategico per la cooperazione europea nel settore dell'istruzione e della formazione* è disponibile online all'indirizzo <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2009:119:0002:01:it:HTML> (ultima consultazione: febbraio 2011).

Il Quadro, al contempo, presenta una definizione degli obiettivi e dei metodi di insegnamento e apprendimento delle lingue, allo scopo di facilitare lo scambio di informazioni tra coloro che sono direttamente coinvolti nell'insegnamento delle lingue straniere e nella valutazione:

“Il Quadro comune europeo di riferimento ha lo scopo di aiutare le persone che operano professionalmente nel campo delle lingue moderne a superare le difficoltà di comunicazione che la diversità dei sistemi scolastici europei fa insorgere. Fornisce ad amministratori scolastici, estensori di programmi, insegnanti e loro formatori, organismi esaminatori, ecc. i mezzi per riflettere sulla loro pratica abituale, così da coordinarsi per rispondere ai bisogni reali degli apprendenti di cui sono responsabili.” (QCER, 2002:1)

Il documento fornisce, nell'ambito dell'educazione linguistica, indicazioni relative a programmi di studio comuni, metodi di insegnamento e valutazione, condivisi a livello europeo. In tal senso il Quadro ha una portata sovranazionale e normativa, poiché si pone come guida per tutti coloro che si occupano di lingue straniere all'interno degli Stati membri, incoraggiando la cooperazione tra le istituzioni educative europee.

“Il Quadro comune europeo di riferimento fornisce una base comune in tutta l'Europa per l'elaborazione di programmi, linee guida curricolari, esami, libri di testo per le lingue moderne, ecc. descrive in modo esaustivo ciò che chi studia una lingua deve imparare per agire in modo efficace. La descrizione riguarda anche il contesto culturale nel quale la lingua si situa. Inoltre il Quadro di riferimento definisce i livelli di competenza che permettono di misurare il progresso dell'apprendente ad ogni stadio del percorso, nella prospettiva dell'educazione permanente.” (QCER, 2002:1)

Il Quadro definisce un repertorio di parametri, categorie e scale a cui gli interessati possono attingere. Tuttavia lo scopo non è quello di prescrivere metodi e programmi indiscutibili, disposti in un sistema chiuso di istruzioni e direttive. Il documento si pone bensì come una serie di opzioni che sboccano in una riflessione aperta e flessibile alle diverse tendenze metodologiche e

glottodidattiche.

1.3.1 Obiettivi, criteri e approcci

Nonostante si tratti di un Quadro di riferimento “non dogmatico”, aperto e flessibile, esso si attiene alla politica dell’UE, la quale mira alla mobilità, all’integrazione e alla cooperazione fra Stati membri. Condizione necessaria per il successo di questa politica europea appare il potenziamento dell’educazione linguistica, ovvero del plurilinguismo inteso come:

“[...] la capacità che una persona, come soggetto sociale, ha di usare le lingue per comunicare e di prendere parte a interazioni interculturali, in quanto padroneggia, a livelli diversi, competenze in più lingue ed esperienze in più culture.”
(QCER, 2002:205)

La capacità plurilingue definita dal Quadro si differenzia dalla capacità multilingue (conoscenza di più lingue da parte di un individuo o coesistenza di diverse lingue in una determinata società), poiché l’individuo che viene a contatto con lingue e culture differenti svolge un’operazione di integrazione delle nuove conoscenze, piuttosto che di classificazione in compartimenti mentali separati. Le lingue non vengono memorizzate dall’individuo in maniera isolata le une dalle altre, senza alcun legame tra di esse, ma possono stabilire rapporti reciproci (interdipendenza linguistica: raggiunta una sufficiente competenza in entrambe le lingue, ciò che viene appreso attraverso una lingua è trasferibile nell’altra lingua²²).

L’educazione linguistica, dunque, non si basa più necessariamente sul raggiungimento della “padronanza” nella lingua straniera considerata singolarmente. L’obiettivo è di conseguire un repertorio plurilinguistico che

²² Si veda la teoria dell’*iceberg* in Cummins, 1979.

consenta a un individuo di rapportarsi con gli altri. Il “Livello Soglia” è il grado minimo di competenza comunicativa necessario per stabilire rapporti relazionali. Raggiungere una competenza parziale (Livello Soglia) non vuol dire rinunciare a una competenza totale, ma conseguire provvisoriamente una competenza che consenta il raggiungimento di un risultato specifico, e arricchirla nel tempo, lungo tutto l’arco della vita.

È evidente che il repertorio linguistico si sviluppa inizialmente in ambito scolastico, per essere poi arricchito progressivamente durante l’arco della vita. Il Quadro si fa portavoce dell’esigenza di una formazione costante e sottolinea l’importanza che hanno le istituzioni scolastiche nel far sì che l’apprendente sviluppi motivazione e fiducia per affrontare autonomamente nuove situazioni linguistiche (imparare a imparare) al di fuori dell’ambiente scolastico.

L’attenzione si sposta sul soggetto apprendente, non più *tabula rasa* su cui il docente riversa il proprio sapere, bensì attore sociale. Il docente deve favorire lo sviluppo di un buon dialogo, affinché l’apprendente possa prima di tutto esprimersi in classe, inoltre deve preparare l’apprendente a utilizzare la lingua per comunicare in situazioni reali fuori dalla classe.

Oltretutto il docente deve esaminare le variabili individuali (età, stili cognitivi, attitudine allo studio, predisposizione per la lingua, ecc.) degli apprendenti e tener conto dei loro bisogni, dei loro stili di apprendimento, stimolarli e motivarli nell’apprendimento delle lingue. Si considerano di conseguenza, in aggiunta alla conoscenza linguistica (grammaticale, fonologica, lessicale, ecc.), anche le risorse cognitive, culturali e relazionali possedute dal soggetto apprendente, il quale deve “sapere la lingua” (abilità linguistiche), “saper fare lingua” (capacità di padroneggiare le abilità linguistiche), “sapere integrare la lingua con i linguaggi non verbali” (gestuali, cinestetici, prossemici,

vestemici, ecc.) e infine “saper fare con la lingua” (capacità di utilizzare la lingua come strumento di azione: si tratta della competenza extralinguistica, strategica, culturale, situazionale e socio-pragmatica) allo scopo di comunicare con efficacia.

L’approccio adottato dal Quadro è orientato all’azione, vale a dire che i soggetti “usano” attivamente la lingua per svolgere compiti di vario genere (linguistico e non). Tali soggetti sviluppano un’attitudine a comunicare grazie a una molteplicità di compiti utili, proposti dal docente, atti a suscitare nell’apprendente il desiderio di imparare la lingua, utilizzando le proprie competenze e risorse in maniera strategica, allo scopo di raggiungere l’obiettivo prefissato. Le strategie adottate da ogni individuo variano in base alle capacità e esperienze diverse di apprendimento di cui ciascuno dispone. L’apprendente deve altresì essere incoraggiato ad assumersi la responsabilità del suo apprendimento, in altre parole a prendere iniziative, a organizzare il proprio apprendimento utilizzando tecniche adeguate, a riconoscere i suoi bisogni, obiettivi, difficoltà e progressi, e ad autovalutarsi.

L’approccio orientato all’azione si fonda sulla capacità di usare la lingua in maniera appropriata per comunicare efficacemente. La lingua è considerata pertanto uno strumento di comunicazione, utile al fine di relazionarsi con gli altri. La forma della lingua lascia spazio all’uso linguistico, e l’attenzione per il soggetto e il suo agire per comunicare divengono predominanti rispetto all’attenzione per la struttura della lingua da insegnare.

In questa ottica di interazione è ritenuto di fondamentale importanza anche il concetto di cultura, in quanto non si può comunicare in modo efficace se non si conosce la cultura a cui la lingua è legata. Il soggetto può sviluppare una sensibilità interculturale solo attraverso i contatti con la cultura della lingua che apprende (ricordando comunque che la lingua è già di per sé strumento di

trasmissione della cultura).

Dal punto di vista educativo si dovrebbe puntare alla formazione globale dell'apprendente, contribuendo alla costruzione di una identità linguistica e culturale, dell'essere per se stessi e dell'essere con gli altri. Il processo educativo-formativo dell'apprendente dovrebbe avvenire mediante tre tappe, le quali consentirebbero di integrare esperienze personali ed esperienze di alterità: l'autopromozione (rapporto con se stessi), la socializzazione (rapporto con gli altri) e la culturizzazione (rapporto con il mondo).

Tenendo conto dell'intento del Quadro, ossia di superare le barriere linguistiche e culturali tra gli Stati membri, il concetto di educazione linguistica prevede l'integrazione linguistica e culturale attraverso il contatto sia con le lingue che con le culture "altre". *"La [cultura] caratterizza e modifica [...] la comunicazione"* (Balboni, 2002:64), che spesso non può avvenire se non si possiede una competenza socio-culturale, se non si conoscono gli atteggiamenti da assumere per avvicinarsi al parlante nativo e gli strumenti da adoperare per interagire linguisticamente con la cultura con cui si vuole entrare in contatto.

"In un approccio interculturale, un obiettivo cruciale dell'educazione linguistica consiste nella promozione dello sviluppo completo e armonioso della personalità dell'apprendente e del riconoscimento della propria identità, arricchita dall'esperienza dell'alterità che si esprime attraverso la lingua e la cultura." (QCER, 2002:2)

In sintesi gli scopi educativi principali del Quadro, esplicitati sin dalle prime pagine del documento, risiedono in primo luogo nel fornire uno strumento ausiliare alle persone interessate nel campo delle lingue affinché possano riflettere sui meccanismi linguistici ed extralinguistici coinvolti nell'acquisizione linguistica, sui principi che stanno alla base della costruzione di un sistema di apprendimento autonomo e continuo delle lingue, sui bisogni, sulle motivazioni e

sui progressi dell'apprendente, sugli obiettivi, sui metodi, sui materiali, sulle risorse e sugli strumenti valutativi a disposizione del docente, sulle possibili agevolazioni per il proprio e l'altrui apprendimento di una lingua. In secondo luogo esso mira a facilitare la comunicazione e lo scambio di informazioni tra coloro che si occupano professionalmente delle lingue, realizzando una batteria uniformatrice di concetti ed espressioni sui quali discutere, oltre a un insieme di scale e di griglie dalle quali trarre spunto.

Per quel che riguarda i criteri per il conseguimento di tali scopi il Quadro si prefigge di essere esaustivo, trasparente e coerente: esaustivo perché dovrebbe poter fornire al lettore un'ampia gamma di possibili obiettivi, metodi, conoscenze, abilità, materiali, punti di riferimento per la misurazione dei progressi e risultati conseguiti; trasparente perché il suo contenuto dovrebbe essere comprensibile e chiaro a tutti gli interessati; coerente perché all'interno del testo non dovrebbero emergere discordanze.

Questi criteri, come sottolineato nel Quadro stesso, non sottintendono l'adozione di un unico approccio dogmatico all'interno del Quadro, bensì abbracciano un'ottica propositiva. Esso accoglie vari punti di vista e metodologie, presentandosi come un documento imparziale (non schierato con un'esclusiva teoria o pratica educativa e linguistica), versatile (adattabile ai diversi contesti educativi), discutibile e dinamico (modificabile in base a perfezionamenti o aggiornamenti) e polifunzionale (utilizzabile per scopi differenti).

1.3.2 La competenza comunicativa

La "competenza comunicativa" consiste nella capacità di comunicare e interagire efficacemente in contesti sociali e culturali diversi.

“Questa competenza non consiste nella sovrapposizione o nella giustapposizione di competenze distinte, ma è piuttosto una competenza complessa o addirittura composita su cui il parlante può basarsi.”

(QCER, 2002:205)

Le competenze acquisite nel corso della propria formazione, alle quali ricorre chi usa una lingua per comunicare o per svolgere compiti e attività richieste nelle diverse situazioni comunicative, possono essere considerate tutte componenti della competenza comunicativa.

Il concetto di competenza comunicativa supera la ristretta nozione di “competenza linguistica”²³, integrandola in una più ampia considerazione: si tratta di una competenza complessa e composita, al cui interno si possono distinguere competenze generali e competenze linguistico-comunicative.

Le “competenze generali” sono quelle di cui ci si avvale per tutte le tipologie di azioni (non necessariamente relative al linguaggio) e comprendono il *sapere, saper fare, saper essere e saper apprendere*.

Il “sapere” consiste nella conoscenza dichiarativa, intesa come conoscenza empirica (sapere ricavato dall’esperienza) o come conoscenza ottenuta attraverso un apprendimento formale (sapere accademico). Si tratta della conoscenza del mondo, strettamente legata alla lingua, poiché ciò che ci circonda possiede un nome e un significato, condiviso da tutti gli individui. Gli elementi della rappresentazione che questi soggetti hanno del mondo si sviluppano già durante l’infanzia, e accrescono con la formazione e l’esperienza acquisite durante il corso di tutta la vita. È evidente che la conoscenza dei fenomeni naturali è distinta linguisticamente in base alla rilevanza che tali fenomeni hanno nel luogo in cui si manifestano. Ciò significa che a contatto con

²³ Secondo la visione chomskiana la competenza linguistica è la conoscenza di un insieme di principi e di regole grammaticali che governano tutte le lingue umane e che tutti gli individui possiedono in maniera innata.

luoghi e culture differenti (competenza socioculturale e interculturale) si acquisiscono nuove conoscenze del mondo. Le conoscenze del mondo appena acquisite non si aggiungono semplicemente alle conoscenze preesistenti, ma ne comportano una ristrutturazione.

Il “saper fare” consiste nell’abilità pratica, ossia nell’esecuzione di un insieme di procedure per lo svolgimento di un’azione. Il saper fare come abilità pratica comprende le abilità sociali (capacità di agire in una determinata circostanza rispettando le convenzioni), le abilità correnti (capacità di svolgere gesti di vita quotidiana), le abilità tecniche e professionali e le abilità del tempo libero. Nella categoria va inserita inoltre l’abilità interculturale (capacità di mettere in rapporto la cultura d’origine con quella straniera).

Il “saper essere” è la competenza esistenziale, ovvero l’insieme delle caratteristiche individuali, degli atteggiamenti, degli stili cognitivi e dei fattori della personalità di un individuo. La competenza esistenziale può essere modificata mediante l’apprendimento di lingue straniere o attraverso l’interazione con lingue e culture “altre” (relativismo culturale).

Il “saper apprendere” consiste nella capacità, prendendo parte a nuove esperienze, di integrare nuove conoscenze con quelle preesistenti. La capacità di imparare, particolarmente importante per l’apprendimento delle lingue, è determinata dagli stili di apprendimento e dai profili dell’apprendente. Le diverse capacità di apprendimento conducono l’apprendente a utilizzare differientemente le strategie a sua disposizione. Il saper apprendere comprende la sensibilità alla lingua e alla comunicazione, la consapevolezza e le abilità fonetiche generali, l’abilità di studio e le abilità euristiche.

Le “competenze linguistico-comunicative” permettono al soggetto di compiere azioni inevitabilmente attraverso l’uso del linguaggio e si suddividono in

competenza linguistica, competenza socio-linguistica e competenza pragmatica.

La “competenza linguistica” è la “*conoscenza e capacità d’uso di strumenti formali con cui si possono comporre e formulare messaggi ben strutturati e dotati di significato*” (QCER, 2002:134). La competenza linguistica comprende il lessico, la fonologia e la sintassi.

La “competenza socio-linguistica” riguarda le conoscenze relative alla dimensione sociale dell’uso della lingua. Ogni comunicazione linguistica tra soggetti di culture differenti viene condizionata dai fattori socioculturali (regole di cortesia, convenzioni e proverbi) e sociolinguistici (differenze di registro e marcatori linguistici).

La “competenza pragmatica” afferisce all’uso appropriato delle funzioni e degli atti linguistici. Si suddivide in competenza discorsiva (capacità di strutturare frasi coerenti), competenza funzionale (capacità di reggere un discorso mantenendo una corretta interazione) e competenza di pianificazione (capacità di organizzare il messaggio secondo schemi testuali condivisi).

*“L’uso della lingua, incluso il suo apprendimento, comprende le azioni compiute da persone che, in quanto individui e attori sociali, sviluppano una gamma di **competenze** sia **generali**, sia nello specifico, **linguistico-comunicative**. Gli individui utilizzano le proprie competenze in **contesti** e **condizioni** differenti e con **vincoli** diversi per realizzare delle **attività linguistiche**. Queste implicano i **processi linguistici** di produrre e/o di ricevere **testi** su determinati **temi** in **domini** specifici, con l’attivazione delle **strategie** che sembrano più adatte a portare a buon fine i **compiti** previsti. Il controllo che gli interlocutori esercitano su queste azioni li porta a rafforzare e a modificare le proprie competenze.”* (QCER, 2002:12)²⁴

Le competenze linguistico-comunicative dell’individuo vengono utilizzate per le abilità linguistiche, le quali possono implicare processi linguistici di produzione e ricezione di testi orali e/o scritti, interazione e mediazione. Tali processi possono avvenire in differenti domini della vita sociale, laddove per

²⁴ Il grassetto è nel testo originale.

dominio s'intende la sfera d'azione, che può essere, secondo la suddivisione del Quadro, pubblica, personale, professionale o educativa (tab. n. 2 in appendice). In ciascun dominio si possono presentare diversi contesti situazionali (in termini di luoghi e momenti, persone coinvolte, oggetti presenti) e diverse condizioni (fisiche, sociali e temporali). Questi processi attivano strategie di volta in volta adatte a soddisfare le esigenze di comunicazione per lo svolgimento di un dato compito e il raggiungimento dello scopo comunicativo.

1.3.3 Livelli comuni di riferimento

La distinzione tra i diversi tipi di competenze consente la comprensione dei "livelli comuni di riferimento", strumento che permette di accertare il progresso degli apprendenti e individuarne, a seconda dei casi, lo stadio di apprendimento e le competenze acquisite.

Il Quadro delinea tre fasce di competenza generale: elementare, intermedio e avanzato (fig. n. 1.3). Ogni fascia è, a sua volta, costituita da due suddivisioni, fino all'individuazione complessiva di sei livelli specifici: "livello di contatto" (*Breakthrough*), "livello di sopravvivenza" (*Waystage*), "livello soglia" (*Threshold Level*), "livello progresso" (*Vantage*), "livello dell'efficacia" (*Effective Operational Proficiency*), "livello della padronanza" (*Mastery*).

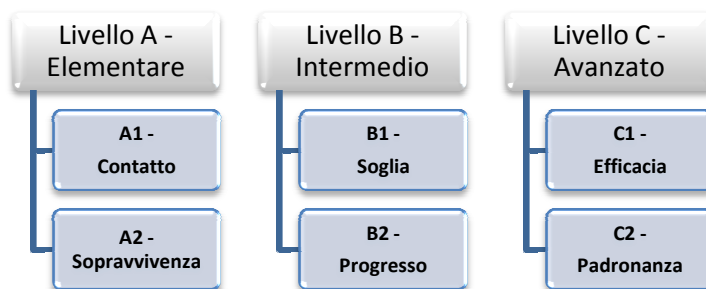


Figura n. 1.3 – Livelli comuni di riferimento

Il Quadro riassume il sistema dei livelli in maniera olistica, attraverso uno schema tassonomico che rappresenta i livelli in una “scala globale” (tab. n. 1.1).

Per ogni livello viene fornito un descrittore che definisce in modo chiaro, breve e semplice che cosa un apprendente è capace di fare con la lingua in quello specifico livello.

Nelle formulazioni di ciascun livello l'apprendente troverà la descrizione di ciò che è in grado di fare, purché dia per scontato che al raggiungimento del livello superiore si riesca in ciò che era stato esposto già al livello inferiore. Ciò vuol dire che gli elementi già citati nei livelli precedenti non sono ripetuti nei livelli successivi, e avanzando verso l'alto si introducono progressivamente nuovi elementi.

Livello avanzato	C2	È in grado di comprendere senza sforzo praticamente tutto ciò che ascolta o legge. Sa riassumere informazioni tratte da diverse fonti, orali e scritte, ristrutturando in un testo coerente le argomentazioni e le parti informative. Si esprime spontaneamente, in modo molto scorrevole e preciso e rende distintamente sottili sfumature di significato anche in situazioni piuttosto complesse.
	C1	È in grado di comprendere un'ampia gamma di testi complessi e piuttosto lunghi e ne sa ricavare anche il significato implicito. Si esprime in modo scorrevole e spontaneo, senza un eccessivo sforzo per cercare le parole. Usa la lingua in modo flessibile ed efficace per scopi sociali, accademici e professionali. Sa produrre testi chiari, ben strutturati e articolati su argomenti complessi, mostrando di saper controllare le strutture discorsive, i connettivi e i meccanismi di coesione.
Livello intermedio	B2	È in grado di comprendere le idee fondamentali di testi complessi su argomenti sia concreti sia astratti, comprese le discussioni tecniche nel proprio settore di specializzazione. È in grado di interagire con relativa scioltezza e spontaneità, tanto che l'interazione con un parlante nativo si sviluppa senza eccessiva fatica e tensione. Sa produrre testi chiari e articolati su un'ampia gamma di argomenti e esprimere un'opinione su un argomento d'attualità, esponendo i pro e i contro delle diverse opzioni.
	B1	È in grado di comprendere i punti essenziali di messaggi chiari in lingua standard su argomenti familiari che affronta normalmente al lavoro, a scuola, nel tempo libero, ecc. Se la cava in molte situazioni che si possono presentare viaggiando in una regione dove si parla la lingua in questione. Sa produrre testi semplici e coerenti su argomenti che gli siano familiari o siano di suo interesse. È in grado di descrivere esperienze e avvenimenti, sogni, speranze, ambizioni, di esporre brevemente ragioni e dare spiegazioni su opinioni e progetti.
Livello elementare	A2	Riesce a comprendere frasi isolate ed espressioni di uso frequente relative ad ambiti di immediata rilevanza (ad es. informazioni di base sulla persona e sulla famiglia, acquisti, geografia locale, lavoro). Riesce a comunicare in attività semplici e di <i>routine</i> che richiedono solo uno scambio di informazioni semplice e diretto su argomenti familiari e abituali. Riesce a descrivere in termini semplici aspetti del proprio vissuto e del proprio ambiente ed elementi che si riferiscono a bisogni immediati.
	A1	Riesce a comprendere e utilizzare espressioni familiari di uso quotidiano e formule molto comuni per soddisfare bisogni di tipo concreto. Sa presentare se stesso/a e altri ed è in grado di porre domande su dati personali e rispondere a domande analoghe (il luogo dove abita, le persone che conosce, le cose che possiede). È in grado di interagire in modo semplice purché l'interlocutore parli lentamente e chiaramente e sia disposto a collaborare.

Tabella n. 1.1 - Livelli comuni di riferimento: scala globale (QCER, 2002:32)

La scala, che procede dall'alto (livello di padronanza) verso il basso (livello di contatto), è uno strumento che permette di rilevare il progresso degli apprendenti durante il processo di acquisizione linguistica, attraverso la misurazione delle competenze acquisite in base ai parametri del modello.

Il sistema dei livelli si presenta come una struttura ad “albero flessibile”, giacché l'insieme dei descrittori può essere scomposto in ulteriori sottocategorie. Il Quadro suggerisce la possibilità di delineare tre “livelli potenziati”: un livello di *sopravvivenza potenziato* (A2+) posto tra il livello A2 e B1, un livello *soglia potenziato* (B1+) posto tra il livello B1 e B2, un livello *progresso potenziato* (B2+) posto tra il livello B2 e C1. Ognuno, per rispondere ai bisogni locali, può sviluppare diramazioni supplementari o ridurre le ramificazioni esistenti in diversi punti (ad es. A2.1, A2.2, ecc.), l'importante è mantenere il riferimento a un sistema comune.

In ogni caso occorre la massima cautela nell'utilizzo della scala. I livelli riflettono infatti una struttura verticale graduata, che però non tiene conto della dimensione orizzontale. L'apprendente può fare progressi laterali, ovvero, piuttosto che accrescere la conoscenza in una stessa categoria avanzando verso la stessa direzione, può fare progressi ampliando la gamma delle abilità e delle conoscenze linguistiche. Progredire vuol dire far incrociare le dimensioni orizzontale e verticale: salire lungo la scala verticale e compiere una serie più ampia di attività dello stesso livello. Anche se sulla scala i livelli sembrano essere equidistanti, questo non vuol dire che gli apprendenti impiegano lo stesso tempo per raggiungere i diversi gradi di competenza. Qualche esperto ha proposto l'immagine diagramma “a cono di gelato” per rappresentare i livelli. Il cono tridimensionale infatti si allarga verso l'alto, descrivendo in modo più soddisfacente le due dimensioni della scala.

Il modello a sei livelli gode di un ampio consenso, si è ispirato alle certificazioni già esistenti elaborate da vari organismi (tra i quali l'ALTE - l'Associazione degli Esaminatori di Lingua in Europa), e attualmente funge da "strumento di taratura" oggettivo e condiviso per numerosi organismi ed enti certificatori.

I livelli di competenza sono orientati, oltre che all'esaminatore per la collocazione nei livelli e per l'attribuzione di punteggi, all'autore per la costruzione delle prove e all'utente per comprendere il proprio livello di competenza. Pertanto i descrittori sono espressi in termini positivi già a partire dai livelli più bassi, poiché l'apprendente, se si avvale dei descrittori come riferimento per fissare i propri obiettivi, troverà più utile e stimolante avere la descrizione di "ciò che è in grado di fare".

Per consentire all'apprendente di valutare cosa è in grado di fare a ogni livello proposto, il Quadro fornisce, in aggiunta alla scala globale, una scheda per l'autovalutazione (tab. n. 1 in appendice) basata sui sei livelli, distinta inoltre secondo le quattro abilità linguistiche. La suddivisione nelle quattro abilità linguistiche presenta una novità per quanto riguarda il parlato, che viene ulteriormente suddiviso in interazione (conversazione) e produzione orale (monologo). La griglia di autovalutazione contiene descrittori analitici per consentire all'apprendente di valutare autonomamente il proprio stadio di competenza.

1.3.4 Il QCER: criteri, principi e tipologie di valutazione

Il tema della valutazione è trattato dal Quadro in maniera estesa ed esauriente. L'apporto più significativo è costituito proprio dall'introduzione dei

livelli di competenza. I livelli di competenza e i relativi descrittori costituiscono una risorsa proficua per la valutazione poiché favoriscono la formulazione di criteri di giudizio e permettono la comparazione delle certificazioni esistenti.

Tramite criteri di riferimento è possibile commisurare il raggiungimento di un determinato obiettivo, che può essere definito in termini di competenza (generali o linguistico-comunicative) o in termini di abilità (ricezione, produzione, interazione e mediazione). Mediante le scale dei descrittori delle abilità comunicative si fornisce una descrizione pratica della capacità linguistica sviluppata dall'apprendente. Possono essere utilizzate per l'elaborazione delle prove, per comunicare il resoconto dei risultati in maniera più concreta e realistica, per la valutazione (continua e/o sommativa) dell'insegnante o per l'autovalutazione. Questo tipo di descrittori sembra funzionare in maniera efficace perché descrive "che cosa" l'apprendente è in grado di fare e "come" è in grado di farlo, agevolandone l'interpretazione. Attraverso le scale dei descrittori di singole competenze si provvede a una descrizione olistica. Tali scale possono essere usate per la valutazione dell'insegnante, per l'autovalutazione e per la valutazione della prestazione da parte dell'esaminatore (in tal caso i descrittori, da impressioni personali, vengono trasformati in giudizi circostanziati, dando vita a un quadro uniformato e condiviso da più valutatori).

I descrittori possono presentarsi sotto forma di scala (i descrittori di categorie differenti sono raggruppati in un'unica area per ogni singolo livello in modo olistico), di liste di controllo (i descrittori sono raggruppati sotto competenze parziali e particolareggiate per ogni singolo livello in modo analitico) o di griglia (si tratta di uno schema analitico a doppia entrata, un insieme di scale parallele distinte in verticale per livello e in orizzontale per categoria). Le griglie a loro volta si possono distinguere in scala di competenza (la griglia contiene la selezione dei

livelli pertinenti per determinate categorie) e in scala di classificazione (la griglia contiene per ogni categoria un descrittore che precisa lo *standard* di accettabilità).

Ogni istituzione e/o nazione possiede attualmente una valutazione dei risultati peculiare (che può esprimersi in numeri, ad es. su una scala che va da 1 a 10, oppure in giudizi, ad es. ottimo, buono e sufficiente, ammesso e non ammesso, ecc.). È importante, nell’ottica di una standardizzazione a livello europeo e internazionale, poter stabilire una relazione tra la valutazione dei risultati e i livelli di competenza che renda più trasparente la relazione che intercorre tra gli esami in lingue europee differenti. Tramite le scale di descrittori è possibile comparare certificazioni esistenti tra istituzioni e tra nazioni diverse.

Il Quadro propone una scala di competenze sulla quale poter localizzare i diversi esami.

9	C2
8	C1
7	B2+
6	B2
5	B1+
4	B1
3	A2+
2	A2
1	A1

Tabella n. 1.2 - Scala di competenza (QCER, 2002: 51)²⁵

Questa scala consente una relazione approssimata tra i livelli di competenza e la gamma degli *standard* utilizzati nei vari paesi e/o organizzazioni.

Il Quadro descrive cinque metodi per il confronto di valutazioni diverse: l’equivalenza (produrre versioni diverse della stessa prova), la taratura (uniformare i risultati di test differenti a una scala comune), le procedure

²⁵ La colonna di destra non è presente nel documento originale.

statistiche di controllo (effettuare controlli di qualità dei test tramite la determinazione di coefficienti statistici), la standardizzazione (convertire i risultati dei test in base a degli *standard*) e l'armonizzazione condivisa (raggiungere un accordo attraverso il dialogo).

Si è ancora lontani dal poter equiparare in maniera precisa le classificazioni dei vari risultati, ma il Quadro ha dato il via al cambiamento.

Il Quadro inoltre indica tre principi ai quali ci si deve attenere nella strutturazione di una prova di valutazione: la validità, l'affidabilità e la fattibilità. Si tratta di alcune indicazioni operative che vanno assunte necessariamente per determinare criteri e *standard* di valutazione.

Una procedura di valutazione si può definire valida se è possibile dimostrare che ciò che si fa valutando corrisponde a ciò che si vuole valutare, e dunque l'esito ottenuto riporta con esattezza la competenza del candidato esaminato.

Un test si può ritenere affidabile se permette di classificare i candidati, ai quali viene somministrata la stessa prova in momenti diversi, allo stesso modo.

Una prova deve essere inoltre fattibile, ossia deve essere agevole per i valutatori al momento della correzione.

“Se due diverse organizzazioni o due diverse aree geografiche basano le decisioni relative alla valutazione della stessa abilità su criteri riferibili agli stessi standard, se gli standard stessi sono validi e adatti ai due contesti presi in considerazione e se nelle prove progettate per la valutazione e nell'analisi delle prestazioni gli standard sono applicati in modo coerente, i risultati ottenuti nei due sistemi dovrebbero essere comparabili.

(QCER, 2002:218)

La valutazione è un'attività complessa che può comprendere la definizione e/o comprensione di vari aspetti (obiettivi, criteri, contenuti delle

prove, modi e tempi di verifica, tipologia di giudizio) e avere varie finalità (valutazione del profitto complessivo dell'allievo sia dal punto di vista dei contenuti acquisiti che dal punto di vista educativo-formativo, valutazione delle sole competenze acquisite, valutazione delle competenze già possedute, ecc.). Il Quadro descrive ventisei tipologie di valutazione (ognuna della quali è riunita in coppie generalmente oppostive), ciascuna legata a un contesto educativo determinato o a uno scopo valutativo specifico (tab. n. 3 in appendice).

L'insieme dei criteri, principi, scale e tipologie per la valutazione è concepito dal Quadro come un punto di riferimento, non come uno strumento univoco, concreto e rigido.

Questo atteggiamento spinge il lettore a essere selettivo e critico nei confronti delle indicazioni contenute nel Quadro, preparandosi a fornire possibili integrazioni, modifiche e miglioramenti.

1.3.5 Punti deboli del QCER

È evidente che il Quadro stesso riconosce di non costituire un manuale sistematico e prescrittivo per l'elaborazione e la valutazione di test, ma di fornire linee guida agli interessati del settore.

Sebbene si riconosca al Quadro il grande contributo di aver reso più trasparente la politica europea nel campo dell'apprendimento e della certificazione delle lingue moderne, si possono ovviamente riscontrare punti deboli che andrebbero rivisti sotto alcuni aspetti. Sembra necessario sottolineare l'assenza, tuttora, di approcci modulari che permettano di differenziare i metodi di apprendimento, valutazione e certificazione per il riconoscimento di competenze plurilingue, che tengano conto delle differenze sincroniche (in un dato momento

del percorso di apprendimento rispetto a un altro), delle differenze diacroniche (lungo le tappe che costituiscono il percorso di apprendimento) e delle differenze personali (da un individuo a un altro). Un approccio modulare e multidimensionale permetterebbe di valutare specificamente le competenze, rafforzando la struttura generale del Quadro.

Il carattere generale e onnicomprensivo che lo contraddistingue ha spinto enti certificatori e associazioni a elaborare *a posteriori* integrazioni al Quadro, documenti più specifici e operativi (ad es. *The User Guide for Examiners* pubblicata dall'ALTE).

Tuttavia il Quadro resta sicuramente un'importante punto di riferimento (avendo contribuito a una definizione comune dei livelli di riferimento) per la coordinazione e uniformazione a livello europeo dei contenuti delle prove e dei criteri di valutazione.

Non bisogna dimenticare che lavori successivi rispetto a esso, compiuti e ancora da compiere, sono e potranno essere numerosi.

1.4 Il Portfolio Europeo delle Lingue

Il *Portfolio Europeo delle Lingue* nasce, in correlazione con il QCER, allo scopo di facilitare la realizzazione degli obiettivi, degli *standard* e delle metodologie proposte dal QCER, e fa esplicito riferimento ai sei livelli di competenza linguistica. Si tratta di un documento personale che appartiene all'apprendente e che consente di documentare le proprie competenze e i propri progressi per una vasta gamma di lingue. Esso è composto da tre parti: il Passaporto delle Lingue, la Biografia linguistica e il Dossier (fig. n. 1.4).

Il **Passaporto delle lingue** fornisce una panoramica della *proficiency* dell'apprendente raggiunta in un particolare momento della propria vita, definita in termini di competenze linguistico-comunicative secondo i livelli comuni di riferimento. Esso può includere sia la valutazione dell'insegnante, delle istituzioni scolastiche ed educative, e degli enti esaminatori, attraverso la registrazione delle certificazioni riconosciute ufficialmente sulla base dei sei livelli comuni di competenza linguistica, sia le competenze e le esperienze interculturali certificabili attraverso una griglia di autovalutazione (proposta dal QCER), compilata dall'apprendente stesso e aggiornata regolarmente.

La **Biografia linguistica** contiene un resoconto più dettagliato dell'apprendimento di un individuo. È progettata per facilitare un personale coinvolgimento nella pianificazione del proprio apprendimento. Essa incoraggia un apprendente a riflettere sul proprio progresso, dal momento che lo conduce a esprimere le proprie competenze in una data lingua e a valutare le proprie esperienze linguistiche e culturali, acquisite dentro e fuori i contesti dell'educazione formale, esperienze che altrimenti non potrebbero essere certificate e riconosciute.

Il **Dossier** offre all'apprendente l'opportunità di selezionare e raccogliere materiali rilevanti (certificati, attestati, lavori svolti, ecc.) per documentare e illustrare competenze ed esperienze acquisite, riportate nel Passaporto delle lingue e nella Biografia linguistica.

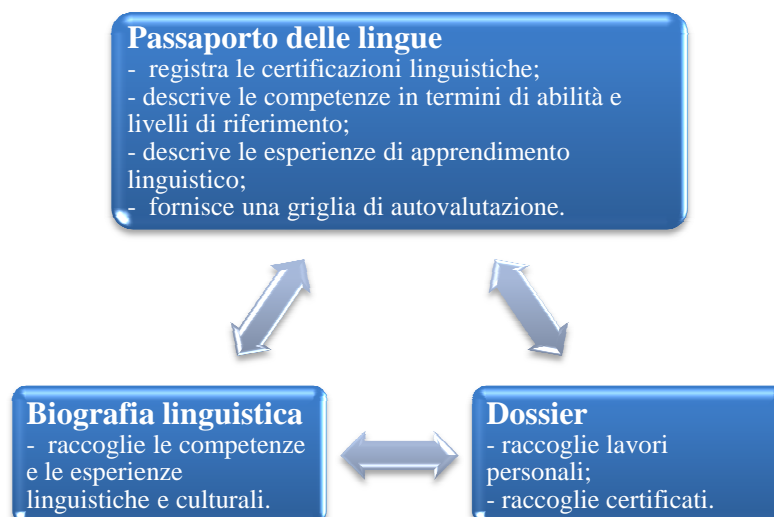


Figura n. 1.4 - P.E.L.

Il Portfolio è un documento approvato dalla Commissione Lingue Moderne del Consiglio d'Europa e costituisce uno *standard* con validità internazionale.

Il documento è stato concepito con lo scopo di esplicitare due funzioni: una pedagogica e una documentaria (illustrate in dettaglio nelle due guide *ELP: Guide for Developers of an European Language Portfolio*²⁶ e *ELP: a Guide for Teachers and Teacher Trainers*²⁷).

Dal punto di vista documentario, il Portfolio mira ad attestare le competenze e le esperienze linguistiche e culturali dell'apprendente in modo complessivo e trasparente. Gli strumenti contenuti nel Portfolio aiutano l'apprendente a determinare in modo autonomo le competenze raggiunte nell'apprendimento di una o più lingue, al fine di poterle illustrare a terzi in modo dettagliato e riconosciuto a livello internazionale.

²⁶ Il documento *ELP: Guide for Developers of an European Language Portfolio* è disponibile online all'indirizzo: http://www.coe.int/T/DG4/Portfolio/documents_intro/Equide.pdf (ultima consultazione: febbraio 2011).

²⁷ Il documento *ELP: a Guide for Teachers and Teacher Trainers* è disponibile online all'indirizzo: http://www.coe.int/T/DG4/Portfolio/documents/ELPgguide_teacherstrainers.pdf (ultima consultazione: febbraio 2011).

Il Portfolio può essere indirizzato a individui che hanno un ruolo decisivo per la carriera scolastica e/o professionale dell'apprendente, i quali sono interessati alla documentazione di diplomi, certificati, valutazioni scolastiche, partecipazioni in programmi di scambio ed esperienze pertinenti per la determinazione delle conoscenze di un apprendente.

Non sempre l'apprendimento linguistico può essere documentato attraverso attestati, pertanto può essere utile avere a disposizione uno strumento che, in modo trasparente e condiviso, esponga competenze linguistiche ed esperienze culturali conseguite nel corso della propria vita. Poiché lo scopo del Portfolio non è quello di sostituire i titoli conseguiti mediante il superamento di esami formali, questo documento si presenta come un'integrazione di tali titoli, fornendo informazioni aggiuntive relative a esperienze informali che hanno influito nell'apprendimento linguistico.

Dal punto di vista pedagogico il documento mira a rendere più consapevole e ponderato il processo di apprendimento di un individuo, agevolando l'autovalutazione dei propri progressi e l'autodeterminazione dei propri bisogni e obiettivi.

La funzione pedagogica del Portfolio è duplice, poiché attraverso il Passaporto l'apprendente può autovalutare sommariamente le proprie competenze e rendersi gradualmente più responsabile del proprio apprendimento. Allo stesso tempo, mediante la selezione dei documenti per il Dossier, egli può autovalutare criticamente la propria formazione, è necessario infatti, affinché il documento sia credibile, che l'apprendente compia un'autovalutazione assumendo un atteggiamento responsabile.

Nella *ELP: a Guide for Teachers and Teacher Trainers* vengono indicati tre elementi basilari del processo di apprendimento: la necessità da parte

dell'apprendente di valutare la propria competenza nel complesso; il bisogno di comprendere il proprio apprendimento a uno stadio specifico; e infine la capacità di svolgere compiti individuali e far fronte autonomamente a obiettivi specifici di apprendimento. L'autovalutazione è un processo che richiede all'apprendente la capacità di riflettere sulle proprie capacità e abilità e di agire in modo indipendente, imparando a comprendere il proprio stile cognitivo. L'autovalutazione è dunque parte integrante del processo di apprendimento in termini di sviluppo dell'autonomia e della capacità di "imparare a imparare" ed è uno degli scopi perseguiti dal Consiglio d'Europa. L'autovalutazione a cui mira il Portfolio non intende sostituire la valutazione degli insegnanti, delle istituzioni scolastiche o degli enti certificatori, giacché la valutazione compiuta da altri individui fornisce un giudizio esterno e oggettivo sulle competenze e abilità dell'apprendente. Il Portfolio concede spazio sia alla autovalutazione che alla registrazione di esami e certificazioni ufficiali, reputandoli l'uno complementare all'altro.

Attraverso lo sviluppo della responsabilità personale e della consapevolezza del proprio apprendimento, il Portfolio può fornire un sostegno ai giovani per accrescere le competenze necessarie per una partecipazione attiva nella società. Esso può essere uno strumento adatto a motivare lo sviluppo del plurilinguismo di un apprendente per tutta la durata della vita, a tutte le età, in contesti differenti, in situazioni educative formali e informali, durante l'istruzione obbligatoria e post-obbligatoria, e a valorizzare lo sviluppo di repertori linguistici che sono adattabili a condizioni e circostanze variabili.

Il Portfolio può essere considerato, in accordo con la politica del Consiglio d'Europa, uno strumento atto a promuovere il rispetto per la diversità linguistica e culturale, la comprensione reciproca e la tolleranza, e al contempo lo sviluppo

dell'autonomia, la diversificazione dell'apprendimento individuale e l'apprendimento permanente.

La *ELP: a Guide for Teachers and Teacher Trainers* (pp. 3-16) delinea due applicazioni pratiche del Portfolio che ne dimostrano l'efficacia nel consolidamento dell'autonomia degli apprendenti più giovani e nel potenziamento dell'apprendimento in età adulta, rispettando così la sua duplice funzione pedagogica.

Con apprendenti giovani l'insegnante può iniziare lavorando sul Dossier (fig. n. 1.5). Gli apprendenti vengono incoraggiati a selezionare e raccogliere i lavori più soddisfacenti e rappresentativi delle proprie competenze. In una fase successiva, attraverso l'introduzione della Biografia, l'insegnante aiuta gli apprendenti a ripercorrere i progressi del loro apprendimento e a fissare nuovi obiettivi. Nell'ultimo stadio l'insegnante introduce il Passaporto, così che gli apprendenti possano autovalutarsi mediante la griglia di autovalutazione proposta dal QCER.

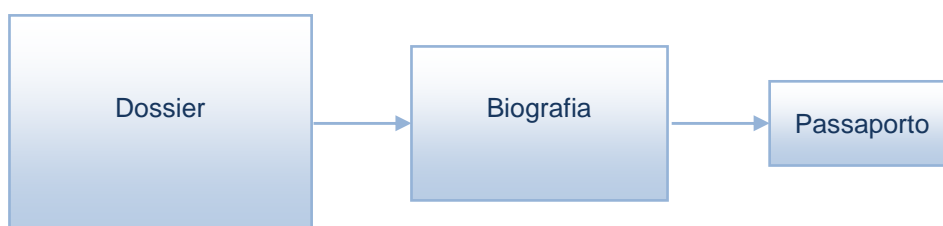


Figura n. 1.5 - Applicazione pratica I (*ELP: a Guide for Teachers and Teacher Trainers*:16)

In tal modo l'insegnante, attraverso il percorso completo di analisi dei tre documenti, conduce l'apprendente adolescente a diventare indipendente nel suo processo di apprendimento.

Con gli adulti il processo potrebbe essere invertito (fig. n. 1.6). Il Passaporto viene introdotto all'inizio per permettere agli apprendenti di riflettere sul proprio grado di competenza. In un secondo stadio viene introdotta la

Biografia al fine di fissare gli obiettivi in vista di un apprendimento futuro e nella terza fase i risultati possono essere raccolti nel Dossier. Questi risultati possono, in una fase successiva, fungere da base per fissare nuove tappe. Il processo può ripetersi più volte, per tornare infine al Passaporto per l'aggiornamento delle ultime competenze acquisite.

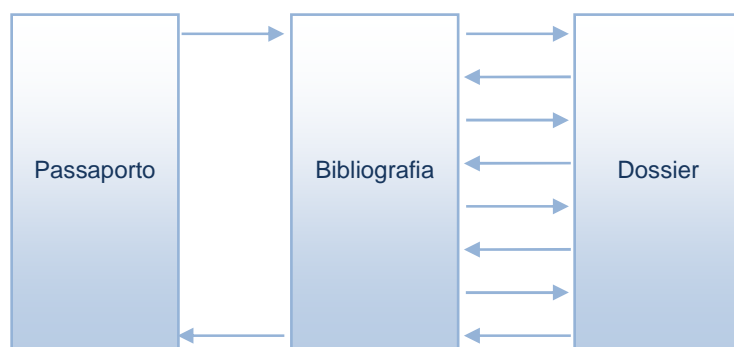


Figura n. 1.6 - Applicazione pratica II (ELP: a Guide for Teachers and Teacher Trainers:16)

In tal modo l'insegnante, attraverso il percorso completo di analisi dei tre documenti, conduce l'apprendente adulto a perseguire l'apprendimento permanente progettandone le varie tappe.

Il Portfolio ha il compito notevole di assicurare una transizione coerente e trasparente, poiché si basa sui livelli comuni di riferimento del QCER ed è riconosciuto a livello internazionale, tra l'istituzione scolastica e l'apprendimento post-scolastico delle lingue.

Il Portfolio è ritenuto di particolare importanza in età adulta, ossia quando diventa più difficile dare prova delle proprie competenze acquisite in ambito extrascolastico e post-scolastico. Per tale ragione il Consiglio d'Europa ha convalidato vari modelli di Portfolio distinti per fascia d'età (per giovani adolescenti tra gli undici e i quindici anni, per ragazzi dai sedici anni fino all'età adulta).

In tal modo si facilita la mobilità internazionale, gli individui possono

spostarsi portando con sé un rendiconto che attesti le competenze possedute, valido indipendentemente dal paese, regione, settore o istituto di origine.

Sebbene il Portfolio contribuisca alla promozione della formazione costante, il Consiglio d'Europa si auspica sempre un più ampio impegno da parte di organismi che si occupano di formazione andragogica per rispondere alla sfida della formazione permanente, continuando ad accettare nuove proposte di modelli di P.E.L. da validare.

1.5 Valutazione

Si è parlato ampiamente fin qui di valutazione attraverso i contenuti riscontrati nel Quadro comune europeo di riferimento per le lingue, ma non è stata ancora presentata una definizione del significato della parola “valutazione”. Secondo il *Dizionario di glottodidattica*, la valutazione:

“È il complesso di operazioni che consiste nel:

- *reperire informazioni sulla quantità e la qualità dell'acquisizione di un allievo (testing);*
- *definire dei parametri (operazione detta anche scaling) da applicare ai dati del test per ottenere un punteggio (operazione detta anche scoring);*
- *elaborazione (a) di un giudizio statistico sul rapporto tra un allievo e il suo gruppo, (b) di un giudizio di merito sull'acquisizione avvenuta e (c) di un giudizio rapportato alla personalità del singolo: i suoi punti di partenza, i suoi progressi, le sue capacità;*
- *esprimere [un] giudizio, che può essere burocratico (il voto in numeri o in lettere) o personalizzato (il giudizio scritto, in cui si fa una diagnosi e, se necessario, si suggerisce una terapia di ripasso, rinforzo o recupero).”²⁸*

²⁸ Cfr. Balboni, 1999, s.v. “valutazione”.

Per prima cosa notiamo la definizione di valutazione in termini di un “complesso di operazioni”, che comprendono la verifica (*testing*), lo *scaling*, lo *scoring* e l’attribuzione di un giudizio.

Spesso i termini “verifica” e “valutazione” vengono utilizzati interscambiabilmente come se fossero sinonimi. In realtà la glottodidattica li distingue. La verifica è una parte del processo di valutazione e consiste nella raccolta quantitativa e qualitativa di dati relativi l’acquisizione di competenze. Si tratta di una valutazione oggettiva degli obiettivi didattici raggiunti. Il conseguimento degli obiettivi educativi rientra invece nel raggio d’azione di un sistema più complesso.

La valutazione si configura come un processo articolato in varie fasi in cui prendono parte numerose variabili. Valutare in senso stretto vuol dire attribuire un valore a qualcosa, dopo aver effettuato un confronto tra i risultati ottenuti e i risultati attesi. L’atto del valutare però non si limita a questa misurazione tra dati empirici e criteri prestabiliti; è un processo in cui i dati oggettivi (esiti del confronto), raccolti attraverso la verifica, vengono giudicati. L’atto del giudicare presuppone un’operazione soggettiva, ossia un’interpretazione personale dei risultati, distribuiti lungo una scala di riferimento presa a modello. Indubbiamente l’interpretazione è influenzata da fattori relativi a obiettivi e finalità della valutazione, strategie e tecniche utilizzate per la definizione dei criteri, e *background* del valutatore.

Il processo di valutazione in senso ampio consta, dunque, di due momenti principali: la rilevazione dei dati (che comprende una serie di procedure volte a raccogliere i dati, a porli su una scala definita *a priori* e a confrontarli) e la valutazione vera e propria (che comprende l’analisi interpretativa dei risultati e l’attribuzione di un giudizio o punteggio).

Questo tipo di valutazione tende a misurare ciò che ci si è prefissati di misurare con la somministrazione di una determinata prova ma non si estende a misurare il processo di apprendimento dell'individuo. Una valutazione più autentica comporterebbe la possibilità di esprimere un giudizio più esteso sul percorso di apprendimento, sullo sviluppo delle capacità metacognitive dell'apprendente, della capacità di *problem solving*, della capacità di autovalutarsi e della capacità di mettere in campo le conoscenze acquisite in situazioni di vita reale. La valutazione mirerebbe così alla verifica degli obiettivi sia didattici sia educativi, abbracciando la complessità dell'essere umano e della lingua come strumento sociale.

Un altro aspetto che il valutatore deve considerare durante un'azione valutativa è la soggettività dell'individuo, il quale, sottoposto a una verifica, può manifestare involontariamente fattori emotivi e fisiopsichici che possono alterare la sua *performance*.

La valutazione a cui si fa riferimento rispecchia l'accezione inglese di "*evaluation*", ossia quel processo che comprende aspetti quali la progettazione di un test; la scelta di metodologie, criteri, obiettivi, finalità; l'attuazione della prova (scelta di tempi, modi e luoghi per la verifica); la misurazione mediante un modello oggettivo definito *a priori*; l'attribuzione di un valore influenzato dalla soggettività del valutatore che tiene conto di fattori relativi all'apprendente; e infine la formulazione di giudizi e decisioni sulla base dei risultati ottenuti. Il processo di *evaluation* coinvolge tutti gli elementi e le fasi di strutturazione e attuazione della valutazione. Si tratta di un'attività rivolta soprattutto a coloro i quali si occupano di ricerca e sviluppo nel campo della valutazione. La valutazione nell'accezione inglese di "*assessment*" è ristretta alla valutazione di profitto, finalizzata cioè a valutare e quantificare con un voto l'acquisizione delle

competenze. Si tratta di un'attività svolta più che altro dagli insegnanti.

Bachman illustra la relazione tra “assessment” e “evaluation” attraverso la rappresentazione che segue:

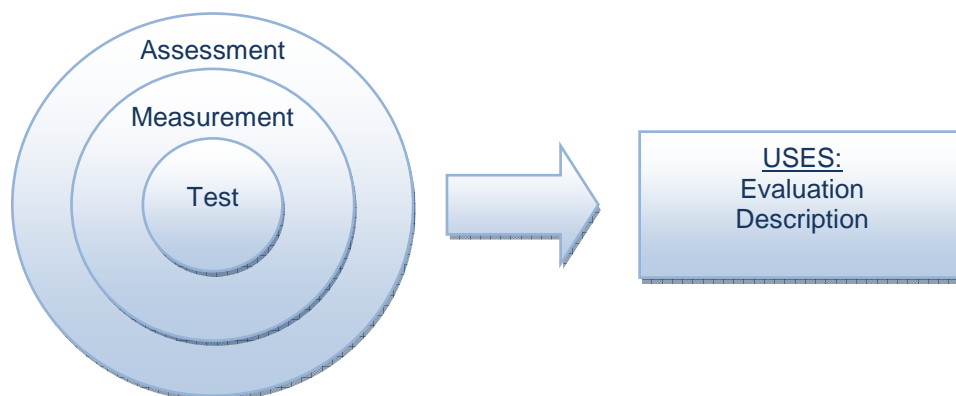


Figura n. 1.7 – Rapporto tra assessment, misurazione, test (Bachman, 2004:10)

La misurazione è un aspetto della valutazione che implica la quantificazione, ossia l'attribuzione di un numero sulla base di norme e procedure prestabilite (criteri, specificazioni, metodi di valutazione e sistemi di votazione). Si differenzia dalla valutazione qualitativa, ovvero dalla formulazione di descrizioni verbali. L'assegnazione di numeri a *performance* osservabili si estende a caratteristiche o abilità inosservabili, non percettibili direttamente.

Il test è un particolare tipo di misurazione progettato per esortare una specifica *performance* o un determinato comportamento, tramite cui è possibile inferire sulle abilità di un soggetto (Carroll, 1968:46).

Con questa rappresentazione Bachman, rifacendosi alla rappresentazione fornita da Lynch (2001), mostra la relazione tra i concetti sovraordinati e subordinati. Pertanto, la valutazione comprende la misurazione quando le decisioni vengono prese sulla base di dati quantitativi, e la misurazione comprende il test quando le decisioni si basano su uno specifico comportamento indotto.

Gli usi possibili di una valutazione possono afferire, conformemente alla raffigurazione appena illustrata (fig. n. 1.7), alla formulazione di giudizi e decisioni in ambito educativo (*evaluation*), oppure a descrizioni, utili in ambito di ricerca, sulle caratteristiche dell'individuo, sul rapporto tra la competenza linguistica e l'uso della lingua, e sui processi di apprendimento.

La docimologia, nata come studio dei criteri di valutazione per l'elaborazione di tecniche di verifica, oggi ha sviluppato una funzione pedagogica, assumendo le vesti di una scienza che studia tutto ciò che è connesso con la misurazione e la valutazione in ambito educativo, dalle procedure di valutazione alle tecniche di misurazione, ai sistemi di votazione, sino al comportamento di esaminatori ed esaminati. Si assume un interesse per l'aspetto qualitativo, oltre che esclusivamente quantitativo, ossia una considerazione per l'intero processo e per le interazioni tra la valutazione nel senso stretto del termine e le caratteristiche dei soggetti coinvolti, le condizioni dell'ambiente, i metodi e gli strumenti didattici. Alla fine ne risulta una commistione di misurazione oggettiva di dati e di interpretazione soggettiva dei risultati.

1.6 Certificazione

Il concetto di certificazione si discosta dal concetto di valutazione. Secondo la definizione del *Dizionario di glottodidattica*, la certificazione è:

*"[una] Attestazione ufficiale, di solito rilasciata da un organismo statale o di alto prestigio glottodidattico, che attesta la conoscenza di una lingua [...]."*²⁹

Certificare vuol dire, dunque, stabilire la conoscenza della lingua e

²⁹ Cfr. Balboni, 1999, s.v. "certificazione".

rilasciare un documento che ne attesti formalmente il possesso. Sostenere un esame di certificazione linguistica vuol dire proporsi a un esame per ricevere un giudizio sul livello di competenza linguistica. La certificazione non presuppone alcun riferimento al percorso formativo, né dal punto di vista della preparazione delle prove, né dal punto di vista del percorso di apprendimento del candidato, dei metodi adottati e dei tempi impiegati. Si tratta di misurare la competenza linguistica in base a criteri stabiliti *a priori*, e valutare ciò che è stato misurato.

In una Europa multilinguistica e multiculturale, caratterizzata dalla promozione della diversità linguistica, della mobilità dei suoi cittadini e della partecipazione attiva nella società, si ha sempre più la necessità di certificare le proprie competenze. Questo bisogno si manifesta su due fronti: l'esigenza di un riconoscimento ufficiale delle competenze di un individuo per la sua autopromozione in campo professionale, educativo, sociale e relazionale, e l'esigenza di trasparenza e comparabilità tra certificati conseguiti in lingue diverse.

La richiesta sempre maggiore di certificazione ha condotto le istituzioni europee a interessarsi di valutazione e certificazione, e definire parametri comuni e condivisi. Grazie alla definizione comune dei livelli di competenza del QCER, ai quali si conformano fedelmente le istituzioni educative e gli enti certificatori, si è attuata un'opera di standardizzazione, che ha contribuito alla riduzione dell'arbitrarietà nell'interpretazione dei livelli di competenza, e ha portato a una maggiore trasparenza e comprensibilità delle certificazioni a livello europeo.

In ambito europeo due associazioni in particolare meritano di essere menzionate per aver contribuito in modo considerevole al livellamento delle diverse interpretazioni relative alle pratiche di valutazione e certificazione: l'EALTA, un'associazione professionale indipendente, finanziata dall'UE, che

vanta della presenza di esperti del settore di verifica e valutazione linguistica; e l'ALTE, un'associazione di istituzioni europee che si occupano di valutazione linguistica.

1.7 EALTA

L'EALTA (*European Association for Language Testing and Assessment*) ha lo scopo di promuovere la conoscenza dei principi teorici del *testing* e della valutazione delle conoscenze linguistiche e di migliorare e condividere le pratiche in materia di verifica e valutazione in Europa.

I principali obiettivi dell'associazione sono:

- condividere competenze professionali nel campo della verifica e della valutazione delle competenze linguistiche;
- aumentare la comprensione pubblica sulla verifica e valutazione delle competenze linguistiche;
- migliorare i sistemi di verifica e valutazione delle competenze linguistiche in Europa;
- sviluppare relazioni con altri gruppi e organizzazioni interessati nel settore della verifica e valutazione delle competenze linguistiche;
- fornire consulenza in fase di verifica e valutazione;
- formare docenti nell'ambito del *testing* e della valutazione.

Per raggiungere tali scopi l'associazione tiene una conferenza ogni anno, organizza seminari e corsi a distanza *Web-Based*, consta di un cospicuo numero di esperti pronti a offrire consulenza a società, istituzioni, organizzazioni governative e non governative e organismi internazionali che operano in materia di valutazione delle competenze linguistiche.

Possono entrar a far parte dell'EALTA sia individui che istituzioni che mostrano un interesse o un legame con il *testing* e la valutazione linguistica.

1.8 ALTE

L'ALTE (*Association of Language Testers in Europe*) è un'associazione di istituzioni europee, ognuna delle quali fornisce esami e certificazioni per la lingua del proprio paese.

La sua costituzione ha origine nel 1990 sotto il nome di *European Economic Interest Grouping* (EEIG), di cui facevano parte inizialmente otto membri. Nasce da un'idea dell'Università di Cambridge e di Salamanca, e attualmente conta 34 membri, in rappresentanza di ventisette lingue, e più di quaranta istituzioni affiliate. Essa include molti organismi fra i più importanti in Europa nel settore della verifica e valutazione della conoscenza delle lingue straniere. L'essere membro a tutti gli effetti dell'associazione è consentito alle sole organizzazioni dei paesi membri dell'Unione europea e dello Spazio Economico Europeo. Tuttavia sono ammessi membri associati, che non hanno però diritto di voto in seno al EEIG. Sia i membri a pieno titolo che i membri associati si occupano di produrre test che certifichino la competenza nella loro lingua madre come lingua seconda o straniera. Le istituzioni affiliate invece sono costituite da coloro che hanno un interesse nel campo del *testing*, ma che non soddisfano i criteri di adesione per appartenervi come membri.

Dal 1990 i membri dell'ALTE lavorano insieme per elaborare uno strumento che descriva e confronti in modo trasparente i propri esami.

I principali obiettivi dell'associazione sono:

- stabilire livelli comuni di competenza linguistica;
- promuovere il riconoscimento transnazionale delle certificazioni in Europa, con riferimento ai livelli comuni di competenza stabiliti dall'ALTE e dal QCER;
- stabilire delle norme comuni per tutti gli stadi del processo di produzione di un test;
- fornire una formazione nel campo del *testing* linguistico;
- sensibilizzare sulle questioni connesse alla valutazione delle competenze linguistiche attraverso meeting e conferenze;
- collaborare a progetti comuni per il miglioramento della valutazione e della verifica delle competenze linguistiche in Europa.

Man mano che le barriere internazionali si dissolvono e accresce la mobilità internazionale nasce l'esigenza di una maggiore agilità nella comunicazione e nella trasferibilità delle qualifiche e certificazioni.

Sin dall'inizio i membri dell'ALTE si sono occupati di definire un quadro di livelli comuni di riferimento e, nel svilupparlo, si sono avvicinati molto al lavoro del Consiglio d'Europa. Il Quadro dell'ALTE comprende sei livelli di competenza linguistica: *Breakthrough*, livello 1, livello 2, livello 3, livello 4, livello 5, equiparati ed equiparabili al QCER (tab. n. 1.3).

Livelli in Quadro ALTE	Livelli in Quadro comune europeo di riferimento per le lingue
ALTE livello 5	Livello C2 – Padronanza (<i>Mastery</i>)
ALTE livello 4	Livello C1 – Efficacia (<i>Effective Operational Proficiency</i>)
ALTE livello 3	Livello B2 – Soglia (<i>Vantage</i>)
ALTE livello 2	Livello B1 – Progresso (<i>Threshold Level</i>)
ALTE livello 1	Livello A2 – Sopravvivenza (<i>Waystage</i>)
<i>Breakthrough</i>	Livello A1 – Contatto (<i>Breakthrough</i>)

Tabella n. 1.3 - Equiparazione dei livelli ALTE e QCER

Il Quadro di riferimento dell'ALTE contiene standard di competenza per ciascuno dei livelli, descritti in modo dettagliato in termini di capacità di sapere fare, e correlabili ai descrittori di competenza del QCER. L'ALTE propone uno schema delle capacità generali distinte per ogni livello:

Livelli	Listening / Speaking	Reading	Writing
C2 Livello 5	CAN advise on or talk about complex or sensitive issues, understanding colloquial references and dealing confidently with hostile questions.	CAN understand documents, correspondence and reports, including the finer points of complex texts.	CAN write letters on any subject and full notes of meetings or seminars with good expression and accuracy.
C1 Livello 4	CAN contribute effectively to meetings and seminars within own area of work or keep up a casual conversation with a good degree of fluency, coping with abstract expressions.	CAN read quickly enough to cope with an academic course, to read the media for information or to understand non-standard correspondence.	CAN prepare/draft professional correspondence, take reasonably accurate notes in meetings or write an essay which shows an ability to communicate.
B2 Livello 3	CAN follow or give a talk on a familiar topic or keep up a conversation on a fairly wide range of topics.	CAN scan texts for relevant information, and understand detailed instructions or advice.	CAN make notes while someone is talking or write a letter including non-standard requests.
B1 Livello 2	CAN express opinions on abstract/cultural matters in a limited way or offer advice within a known area, and understand instructions or public announcements.	CAN understand routine information and articles, and the general meaning of non-routine information within a familiar area.	CAN write letters or make notes on familiar or predictable matters.
A2 Livello 1	CAN express simple opinions or requirements in a familiar context.	CAN understand straightforward information within a known area, such as on products and signs and simple textbooks or reports on familiar matters.	CAN complete forms and write short simple letters or postcards related to personal information.
A1 Livello Break through	CAN understand basic instructions or take part in a basic factual conversation on a predictable topic.	CAN understand basic notices, instructions or information.	CAN complete basic forms, and write notes including times, dates and places.

Tabella n. 1.4 - Overall General Ability³⁰

³⁰ La tabella è consultabile nell'Appendice D del QCER della versione inglese, reperibile online all'indirizzo http://www.coe.int/t/dg4/linguistic/source/Framework_EN.pdf, p. 251 (ultima

È possibile riscontrare un collegamento tra lo schema appena illustrato e la scala globale del QCER. Nello schema dell'ALTE si evince immediatamente la suddivisione per abilità (non presente nel QCER). Dunque il confronto appare costituire un passo in avanti da parte dell'ALTE. Descrizioni più complete e circostanziate sulle competenze sono presenti in ulteriori schemi proposti dall'ALTE, relativi a quattro macro-categorie: studio, lavoro, relazioni sociali e turismo³¹.

Il sistema nel complesso comprende circa 400 indicazioni di capacità, ripartite in 40 categorie (raggruppabili nelle quattro macro-sezioni), che descrivono ciò che è in grado di fare chi usa una lingua. Le descrizioni consistono in una serie di affermazioni (*Can Do statements*) relative alle capacità in una determinata lingua, a un determinato livello, per un determinato argomento, in una delle abilità linguistiche (ascolto/produzione orale, produzione scritta, lettura). Le tavole fornite dall'ALTE chiariscono e approfondiscono i concetti contenuti nel QCER e servono a creare e classificare esami e certificazioni in modo più scrupoloso.

I livelli ALTE permettono un confronto trasversale con le certificazioni ufficialmente riconosciute di numerose lingue europee, per consentire un riconoscimento condiviso del livello di competenza linguistica testata da ogni esame di certificazione, e facilitano la comprensione di ciò che il possessore di un certificato sa fare in una data lingua a un determinato livello.

L'ALTE si propone l'obiettivo di chiarire come certificati conseguiti in lingue diverse possano corrispondere fra loro. Attualmente le certificazioni

consultazione: febbraio 2011).

³¹ Le quattro tabelle sono consultabili nell'Appendice D del QCER della versione inglese, reperibile online all'indirizzo http://www.coe.int/t/dg4/linguistic/source/Framework_EN.pdf, pp. 253-256 (ultima consultazione: febbraio 2011).

linguistiche inserite nel Quadro di riferimento dell'ALTE sono disponibili per le seguenti lingue: basco, catalano, danese, finlandese, francese, greco, inglese, italiano, lussemburghese, norvegese, olandese, portoghese, spagnolo, svedese, tedesco. I livelli e gli esami ALTE sono gli unici ancorati ai livelli del QCER. L'ALTE sta provvedendo all'inserimento di esami prodotti da altri membri e affiliati per le lingue estone, gallese, lettone, lituano, polacco, russo, sloveno, ungherese).

Per la lingua inglese rientrano nel quadro delle certificazioni ALTE gli esami dell'UCLES (*University of Cambridge Local Examination Syndicate*)³². Le corrispondenze tra i livelli del Quadro europeo, i livelli ALTE e le certificazioni dell'UCLES sono illustrate nella tabella seguente:

Livelli QCER	A1	A2	B1	B2	C1	C2
Livelli ALTE	ALTE Breakthrough	ALTE Livello 1	ALTE Livello 2	ALTE Livello 3	ALTE Livello 4	ALTE Livello 5
Esami UCLES	-	Key English Test (KET)	Preliminary English Test (PET)	First Certificate in English (FCE)	Certificate in Advanced English (CAE)	Certificate of Proficiency in English (CPE)

Tabella n. 1.5 - Esami inseriti nel Quadro ALTE per le certificazioni della lingua inglese

Si può osservare, mediante la raffigurazione della lingua inglese, come il sistema di valutazione dell'ALTE abbia una struttura che consenta di delineare con facilità le corrispondenze tra i livelli del Consiglio d'Europa e le certificazioni europee, offrendo al candidato la spendibilità delle proprie certificazioni a livello internazionale.

L'ALTE offre al candidato altresì un'attività di certificazione caratterizzata dall'imparzialità, dalla trasparenza e dalla coerenza, dovuta a procedure ben

³² L'UCLES (*University of Cambridge Local Examinations Syndicate*) nel 2003 si è trasformato in Cambridge ESOL (*English for Speakers of Other Language*) Examinations.

definite per lo sviluppo degli esami stessi condivise da tutti gli enti accreditati, tenuti ad aderire al *Codice Professionale*³³ che ALTE stesso si è imposto.

I membri dell'ALTE offrono la possibilità di sostenere esami standard (*Paper Based Test*) e esami computerizzati (*Computer Based Test*).

L'ALTE, oltre a occuparsi di preparazione di test e rilascio di certificazioni, si occupa di formare esperti della valutazione e di attirare sempre più individui e istituzioni a interessarsi di *testing*. L'ALTE organizza incontri per i suoi membri due volte l'anno, a rotazione, presso le sedi delle istituzioni partecipanti all'ALTE, e ogni tre anni una conferenza internazionale sulla verifica e valutazione delle competenze linguistiche. Inoltre organizza altri eventi a porte aperte per favorire uno scambio di idee e la condivisione delle "pratiche migliori".

In aggiunta, negli ultimi anni l'ALTE ha potuto dare un forte contributo a una serie di progetti del Consiglio d'Europa, tra cui ricordiamo lo sviluppo del Livello Progresso (B2) del QCER e la redazione di *The User Guide for Examiners*, un documento integrativo al QCER.

L'associazione ha partecipato inoltre all'elaborazione del Portfolio EAQUALS (*European Association for Quality Language Services*), uno dei primi portfoli accreditati dal Comitato Validatore del Consiglio d'Europa. Il Portfolio ha avuto l'approvazione dal Consiglio d'Europa nel 2000 e ed è ora disponibile in varie lingue. È stata realizzata recentemente una versione elettronica del Portfolio Linguistico Europeo EAQUALS-ALTE (eELP³⁴).

I membri dell'ALTE partecipano anche alla pubblicazioni di una serie di volumi edita da Milanovic e Weir, insieme all'Università di Cambridge ESOL

³³ Il *Codice professionale dell'ALTE* è disponibile online all'indirizzo: <http://www.alte.org/downloads/index.php?docid=23> (ultima consultazione: febbraio 2011).

³⁴ Il *Portfolio Linguistico Europeo elettronico EAQUALS-ALTE* è disponibile in lingua inglese online all'indirizzo: <http://www.eelp.org/eportfolio/index.html> (ultima consultazione: febbraio 2011).

Examinations e la Cambridge University Press. La serie “*Studies in Language Testing*” affronta una vasta gamma di questioni importanti e nuovi sviluppi nel campo della verifica e della valutazione delle competenze linguistiche ed è un’indispensabile risorsa per candidati, sviluppatori e ricercatori. Tra i più di trenta volumi, si ricordano i volumi 6 e 7, rispettivamente il *Multilingual Glossary of Language Testing Terms* (1998) e il *Dictionary of Language Testing* (1999).

1.9 Riflessioni conclusive

Come emerge da quanto illustrato in questo capitolo sui documenti emanati e sulle iniziative assunte in questi anni, l’educazione plurilinguistica promossa dalle istituzioni educative e dalle organizzazioni comunitarie è certamente un’importante supporto per l’apprendimento delle lingue straniere. La propagazione su larga scala di iniziative e direttive per la promozione del plurilinguismo ha portato a un bisogno urgente di standardizzazione e allo sviluppo di un linguaggio comune, che è sfociato nell’elaborazione del QCER e di altri progetti da parte di enti certificatori.

Il QCER costituisce sicuramente un buon inizio per l’elaborazione di un linguaggio comune sui livelli di competenza e di una metodologia che centra l’attenzione sulla competenza comunicativa in tutte e quattro le abilità, in pieno accordo con la politica dell’UE e con i bisogni della società attuale.

Tuttavia alcuni studiosi della comunità linguistica hanno espresso alcune riserve al QCER. Secondo Weir (direttore del CRELLA, Centro di Ricerca per l’Apprendimento e la Valutazione della Lingua Inglese, presso l’Università del Bedfordshire) il QCER non è uno strumento sufficientemente trasparente e dettagliato per essere considerato un quadro normativo per lo sviluppo di prove

linguistiche (come gli autori stessi del QCER precisano), poiché i progettisti di un test necessitano, non solo dei descrittori sulle capacità possedute da un apprendente a ogni livello, ma anche e soprattutto di specifiche sulle condizioni in cui un apprendente deve mostrare tali capacità e di descrizioni sui criteri di qualità scelti per l'interpretazione dei risultati. Secondo lo studioso è necessario un quadro più esplicito che sia capace di fornire “*evidence of validity*” (il concetto sarà chiarito nel prossimo capitolo).

Anche Linda Taylor (assistente di direzione del *Research and Validation Group* dell'Università di Cambridge ESOL Examination) sostiene che il QCER è un'eccellente punto di partenza ma sono necessarie ulteriori specificazioni per lo sviluppo dei test linguistici.

Per tale motivo i membri dell'ALTE hanno sviluppato un quadro più dettagliato di livelli comuni di riferimento, ispirandosi comunque al QCER, per la valutazione e comparazione dei loro esami.

Il QCER è indiscutibilmente una fonte basilare per l'allineamento dei vari esami ai comuni livelli di riferimento in ambito europeo e il lavoro di enti certificatori quali ALTE e EALTA è senza dubbio fondamentale per la standardizzazione di procedure di progettazione delle prove e di valutazione dei risultati.

Il lavoro delle organizzazioni comunitarie, delle istituzioni educative e degli enti certificatori ha permesso a studiosi e ricercatori nell'ambito del *testing* linguistico di occuparsi con più attenzione e interesse alle questioni di progettazione e valutazione delle prove linguistiche, di cui ci occuperemo nel capitolo successivo.

Testing linguistico e valutazione

2.1 Premesse introduttive

La ricerca di strumenti per la misurazione del profitto ha dato vita agli studi sul “*Language Testing*” (LT), in cui confluiscono sviluppi recenti di varie discipline, tra cui la linguistica, la didattica delle lingue straniere, la docimologia e la statistica metodologica¹. Il *testing* linguistico può essere definito come il complesso dei principi metodologici e delle tecniche che guidano la progettazione, la realizzazione e la somministrazione di prove linguistiche per la verifica dell’apprendimento delle lingue straniere.

Per chiarire l’importanza che oggi ha assunto il *testing* linguistico, si citano le parole di McNamara:

“Testing is a universal feature of social life. Throughout history people have been put to the test to prove their capabilities or to establish their credentials; [...]. In modern societies such tests have proliferated rapidly. Testing purposes of detection or to establish identity has become an accepted part of sport [...], the law [...], medicine [...], and other fields. Tests to see how a person performs particularly in relation to a threshold of performance have become important social institutions and fulfil a gatekeeping function in that they control entry to many important social roles. These include the driving test and a range of tests in education and the workplace. [...] The expertise of those they involved in testing is seen as

¹ La statistica metodologica si occupa dell’analisi e interpretazione dei dati relativi a uno o più fenomeni reali, propri della branca della statistica descrittiva, e dello studio probabilistico e inferenziale dei risultati campionari, argomento della statistica inferenziale.

remote and obscure, and the tests they produce are typically associated in us with feelings of anxiety and powerlessness." (McNamara, 2000:3)

McNamara non si limita alla trattazione del *testing* in campo linguistico, bensì menziona vari campi in cui il *testing*, ossia l'analisi e la verifica delle capacità e competenze, risulta indispensabile sin dai tempi del poema omerico e della leggenda di Arturo. Si preoccupa però di sottolineare che, benché possa sembrare singolare, considerata la sua continua centralità a livello sociale, il *testing* rimane ancorato alla sfera d'azione degli esperti, lontano dalla pratica comune. Anzi se evoca una qualsiasi reazione, probabilmente viene associato ad aspetti ed eventi negativi, come l'immagine di un'aula d'esame in cui viene somministrata una prova cartacea, alla quale si cerca di adempiere lottando contro un intervallo di tempo stabilito per la consegna; oppure l'immagine di una sedia al di fuori di un'aula d'esame occupata da una vittima nervosa in attesa di essere chiamata dagli esaminatori per un colloquio dai toni inquisitori. McNamara afferma che tuttavia "*there is more to language testing than this*" (McNamara, 2000:3).

La natura del *testing* linguistico è cambiata radicalmente nel corso degli anni, si interessa di accertare ciò che un individuo *sa* della lingua, piuttosto che ciò che *non sa*, e ciò che *sa fare* con ciò che sa. Si rinnovano così le forme di valutazione delle competenze linguistiche, dai tradizionali metodi valutativi si passa a forme quali l'osservazione del modo in cui uno studente svolge un compito in classe interagendo con i suoi pari, l'osservazione del modo in cui uno studente svolge un compito al di fuori del contesto scolastico, la sperimentazione di giochi di ruolo in classe, ecc. Si tratta di metodi che tendono all'analisi di una *performance* reale, o di una *performance* che si avvicini a una situazione reale, nella quale lo studente mostra la capacità di saper applicare nella vita reale le conoscenze e le abilità apprese in ambito educativo.

2.2 Presupposti teorici sui modelli di competenza linguistica e suggerimenti alla progettazione di *item*

Gli attuali test linguistici mirano ad appurare l'acquisizione della competenza comunicativa e il grado di possesso delle quattro abilità linguistiche così come sono intese nel QCER. L'attenzione da parte di tutti coloro che sono coinvolti nel campo delle lingue straniere, si è spostata da un'attenzione sulla conoscenza del sistema formale della lingua, a una conoscenza della lingua come strumento di comunicazione utile e necessario. Un tale approccio comunicativo fa riferimento ai reali bisogni comunicativi dell'individuo, in relazione al contesto e agli interlocutori. Di conseguenza, se l'insegnamento tende allo sviluppo della competenza comunicativa, le prove linguistiche devono saper testare tale competenza.

Gli stessi esperti del *testing* linguistico riconoscono di far riferimento per lo sviluppo e l'utilizzo delle prove linguistiche alle teorie sulla competenza linguistica proposte negli ultimi quaranta anni del secolo scorso.

I primi modelli descrittivi atti a definire la competenza linguistica risalgono agli anni Sessanta e contemplano i concetti di abilità e di elementi compositivi della conoscenza linguistica (Lado, 1961; Carroll, 1961). Le abilità linguistiche, suddivise in produzione scritta, produzione orale, comprensione scritta e comprensione orale (*writing, speaking, reading e listening*), sono concepite come distinte dalle componenti della conoscenza linguistica, che a loro volta comprendono grammatica, lessico, fonologia e grafematica.

Lado (1961) focalizza la sua attenzione anzitutto sul sistema grammaticale, sul lessico e sugli aspetti della pronuncia. Questi elementi della conoscenza vengono considerati e analizzati come separati gli uni dagli altri, e

come decontestualizzati da tutto il resto. L'approccio predominante per lo sviluppo delle prove è definito psicometrico-strutturalista ed è caratterizzato da un'enfasi sull'oggettività della valutazione. La forma di *testing* più adatta è conosciuta come "*discrete point testing*", e implica un processo di smembramento delle diverse componenti della lingua al fine di sottoporle isolatamente a una verifica oggettiva. Ciò può essere ottenuto, ad esempio, attraverso il quiz a scelta multipla, un tipo di *item* in cui tanti elementi separati sono diretti a verificare il raggiungimento di un unico obiettivo (le strutture grammaticali, la conoscenza dei vocaboli, ecc.) in maniera oggettiva, poiché ogni domanda ammette, tra una serie di opzioni predeterminate, una sola risposta esatta, alla quale corrisponde un dato punteggio. Gli elementi da testare vengono presentati in un contesto ridotto ai minimi termini, quasi sempre in frasi isolate. Anche le quattro abilità linguistiche, nella prospettiva di Lado, sono considerate in maniera isolata e testate attraverso questa tipologia di prova. I test adottano una visione gerarchica della competenza linguistica in linea con la visione strutturalista della lingua, costruita dal fonema al morfema, alla parola, alla frase. Inoltre, in questa fase, non vengono ancora fornite indicazioni su come le componenti e le abilità linguistiche sono collegate fra loro e su come relazionarle al contesto d'uso.

Nello stesso periodo Carroll (1961), in opposizione al *discrete point test* di Lado, teorizza l'"*integrative test*". Alla base del nuovo orientamento vi è l'idea che le caratteristiche sistematiche della lingua (grammatica, vocabolario e pronuncia) debbano integrarsi fra loro, e abbracciare la comprensione del contesto. L'approccio integrativo considera la lingua nella sua globalità, le prove verificano le competenze linguistiche e le abilità associate tra loro, valutano la conoscenza della lingua autentica e l'effetto comunicativo dell'enunciato. Lo stesso Carroll

afferma che “*If we limit ourselves to testing only one point at a time, more time is ordinarily allowed for reflection than would occur in a normal communication situation, no matter how rapidly the discrete items are presented. For this reason I recommend tests in which there is less attention paid to specific structures points or lexicon than to the total communicative effect of an utterance. [...] The examinee is [...] concerned with [...] the total meaning of the sentence*” (Carroll, 1961:318).

In realtà, però, la costruzione degli *item* sulla base dell'*integrative method* non riflette ancora l'uso della lingua in situazioni di vita reale.

Negli anni Settanta gli studi di linguisti come Halliday (1976), Van Dijk (1977) e Hymes (1972) hanno allargato sempre più il concetto di competenza linguistica, riconoscendo una forte importanza al cotesto (contesto linguistico) e al contesto (contesto situazionale), concetti che permettono di considerare aspetti che vanno al di là del contenuto della singola frase.

Sempre agli inizi degli anni Settanta un altro studioso, Oller (1973), offre un nuovo punto di vista in merito alla lingua e all'uso della lingua. Il suo interesse, più che verso la conoscenza della lingua, è concentrato sui processi psicolinguistici coinvolti nell'uso della lingua. Nell'ottica di Oller l'uso del linguaggio coinvolge due componenti: l'elaborazione del messaggio in tempo reale (produzione o comprensione orale) e la mappatura pragmatica, ossia il ricorrere alla conoscenza del sistema formale della lingua per esprimere e comprendere il significato nel contesto (produzione o comprensione scritta). Le tipologie di *item* (definite, dallo stesso Oller, “*pragmatic tests*”) più adeguate a verificare la capacità di saper integrare la conoscenza grammaticale, lessicale, contestuale e pragmatica sono riconosciute nel *cloze test* e nel dettato. Il *cloze test* è una prova in cui si richiede all'esaminato di completare un testo con le

informazioni mancanti e in cui è sottintesa la capacità di integrare più abilità. Il dettato è una prova atta a stabilire la memoria uditiva, la capacità ortografica e la capacità di comprensione testuale, dunque anch'essa in grado di accertare la capacità linguistica in modo globale.

Gli studi di Hymes (1972) sulla competenza comunicativa hanno esteso l'importanza della lingua come strumento comunicativo, e dunque della capacità di uso della lingua in un contesto, anche agli aspetti socioculturali. Hymes afferma che conoscere una lingua significa andare al di là della conoscenza delle pure regole grammaticali. Vanno considerate anche regole specifiche dal punto di vista culturale che riguardano l'uso della lingua in un dato contesto comunicativo.

L'importanza della teoria di Hymes fu subito percepita, ma impiegò circa un decennio prima di far sentire il suo impatto pratico nelle prove, cosiddette "*communicative language tests*". Si tratta di prove che valutano la *performance*, cioè verifiche nelle quali all'esaminato viene richiesto di assumere un determinato ruolo (elicitando il probabile ruolo sociale che il candidato può avere nel mondo reale) e di dare vita a un atto di comunicazione completo. Il *communicative language test* si discosta dal *pragmatic test* di Oller e rappresenta un profondo mutamento di prospettiva, da una psicologica, che vede la lingua come un fenomeno interno all'individuo, a una sociologica, che si concentra sulle funzioni sociali della lingua vista come un mezzo per relazionarsi con ciò che sta al di fuori dell'individuo.

Negli anni Ottanta l'attenzione verso il concetto di competenza linguistica e gli elementi che intervengono nell'uso e nella conoscenza della lingua hanno continuato a dare vita a interessanti riflessioni.

Canale e Swain (1980) ricordano la relazione tra il messaggio e il

contesto sviluppata negli anni precedenti e contribuiscono a una ridefinizione di competenza linguistica, fondamentale per le ricerche successive in materia di competenza comunicativa e di *testing* linguistico. Canale e Swain ipotizzano quattro elementi della competenza linguistica: grammaticale (relativa al sistema formale della lingua), discorsiva (relativa all'uso della lingua nel contesto in modo coerente e coeso), sociolinguistica (relativa alle regole dell'uso della lingua in termini di ciò che è appropriato in situazioni diverse, per vari scopi e con interlocutori differenti) e strategica (relativa alla capacità di utilizzare tecniche di compensazione, allo scopo di superare un'interruzione di comunicazione causata da un'insufficiente competenza linguistica, che si manifestano attraverso mezzi verbali o manifestazioni non verbali). La competenza strategica, diversamente dalla competenza grammaticale, non è una conoscenza immagazzinata nella memoria dell'individuo, ma una conoscenza che si acquisisce con l'esperienza e consiste nella capacità di sapersi comportare in maniera strategica durante una *performance*. La competenza discorsiva possiede invece sia aspetti della conoscenza intrinseca, che elementi estrinseci che permettono di negoziare il significato in base al contesto.

Le parole di Kramersch (1986), qualche anno dopo, riecheggiano il concetto di competenza linguistica di Canale e Swain:

"[...] interaction always entails negotiating meaning, i.e. adjusting one's speech to the effect one intends to have on the listener. It entails anticipating the listener's response and possible misunderstandings, clarifying one's own and the other's intentions and arriving at the closest possible match between intended, perceived, and anticipated meanings." (Kramersch, 1986:367)

Savignon (1983), pressappoco negli stessi anni, fornisce una più ampia definizione del concetto di comunicazione, riconoscendo l'importanza dei quattro elementi della competenza linguistica sopra descritti:

“It depends on the negotiation of meaning between two or more persons. [...] [It] is context specific. Communication takes place in an infinite variety of situations, and success in a particular role depends on one’s understanding of the context and on prior experience of a similar kind.”

(Savignon, 1983:8-9)

Candlin (1986) describe la competenza comunicativa come:

“the ability to create meaning [...], negotiating the value of convention rather than conforming to established principle. [...] a coming together of organized knowledge structures with a set of procedures for adapting this knowledge to solve new problems of communication that do not have ready-made and tailored solutions.”

(Candlin, 1986:40)

Gli sviluppi sul concetto di competenza comunicativa cercano di offrire una descrizione più completa della conoscenza necessaria per l'utilizzo della lingua, rispetto al precedente quadro delle competenze e abilità definito negli anni Quaranta. Il rinnovato concetto di competenza comunicativa comprende, in aggiunta alla conoscenza delle regole grammaticali, la conoscenza delle modalità di utilizzo della lingua per raggiungere un determinato scopo comunicativo in un dato contesto socioculturale, la dinamicità del processo comunicativo e le variabili relative agli interlocutori che possono intervenire.

Numerosi studiosi, alla luce di queste ultime riflessioni, hanno tentato di tracciare un quadro di riferimento dettagliato per lo sviluppo di prove linguistiche e per l'interpretazione delle *performance*. I test di tipo comunicativi accettano l'ipotesi che la lingua possa essere suddivisa in abilità. Quasi vent'anni dopo, lo stesso Carroll (1980) si discosta dai test di tipo *integrative* e analizza le caratteristiche che una prova deve possedere, indicate nel metodo RACE (*Relevance, Acceptability, Comparability e Economy*), preoccupandosi per la prima volta del punto di vista dell'esaminato. Per accertare con esattezza le competenze di un candidato si deve, secondo quest'ottica, ricorrere a un test che analizzi le abilità separatamente, in circostanze che simulino situazioni reali.

Questa forma di prova è di tipo soggettivo, giacché il valutatore assegna un giudizio sulla base di ciò che osserva, trovandosi a dover interpretare una *performance* piuttosto che dover valutare prove oggettive.

Negli anni Novanta, ulteriori indicazioni sulle differenti componenti della conoscenza inerenti la competenza comunicativa sono fornite da Lyle Bachman (1990²), il quale elabora un modello, detto CLA (*Communicative Language Ability*), che ha fondamenti sia empirici che teorici, e che costituisce ancora oggi una solida base per l'attività di valutazione.

Attualmente il *testing* si presenta sempre più come attività rivolta a comprendere e valutare la *communicative performance* nella sua totalità, cercando di giungere a giudizi conformi alla reale competenza comunicativa, piuttosto che come una serie di prove atte a valutare elementi isolati.

2.3 Il modello di Bachman

Bachman sostiene, alla base dei suoi studi, l'importanza di individuare le abilità che si intendono misurare e i mezzi attraverso i quali si misurano tali abilità, al fine di poter basare su queste osservazioni lo sviluppo e l'utilizzo di prove linguistiche.

Secondo Bachman la capacità di utilizzare la lingua in modo comunicativo implica sia la conoscenza della lingua, sia la capacità di mettere in pratica e

² Il testo di Bachman, *Fundamental Considerations in Language Testing*, riprende il titolo da due opere importanti, pubblicate nello stesso anno: il primo testo sul testing linguistico di Lado (*Language Testing*) e il testo di Carroll (*Fundamental considerations in testing for English language proficiency of foreign students*). Entrambi costituiscono il punto di partenza per i successivi studi nel campo del testing. Lo stesso Bachman asserisce: "In solving my title problem, I also echo my debt to Lado and Carroll; hopefully what I've taken from them is returned in some small measure in the pages that follow." (Bachman, 1990:IX).

usare questa competenza in un contesto comunicativo appropriato.

Nel modello elaborato da Bachman le varie componenti della competenza comunicativa interagiscono le une con le altre, e con il contesto nel quale l'uso della lingua si manifesta.

La competenza linguistico-comunicativa (CLA) comprende tre componenti principali: la "competenza linguistica" (l'insieme delle specifiche componenti della conoscenza utilizzate nella comunicazione), la "competenza strategica" (la capacità razionale di negoziare il significato in un determinato contesto) e i "meccanismi psicofisiologici" (processi neurologici e psicologici coinvolti nell'esecuzione di un messaggio percepito come un fenomeno fisico).

Queste componenti interagiscono con il "contesto situazionale" e con la "conoscenza del mondo", secondo la modalità illustrata nella figura seguente:

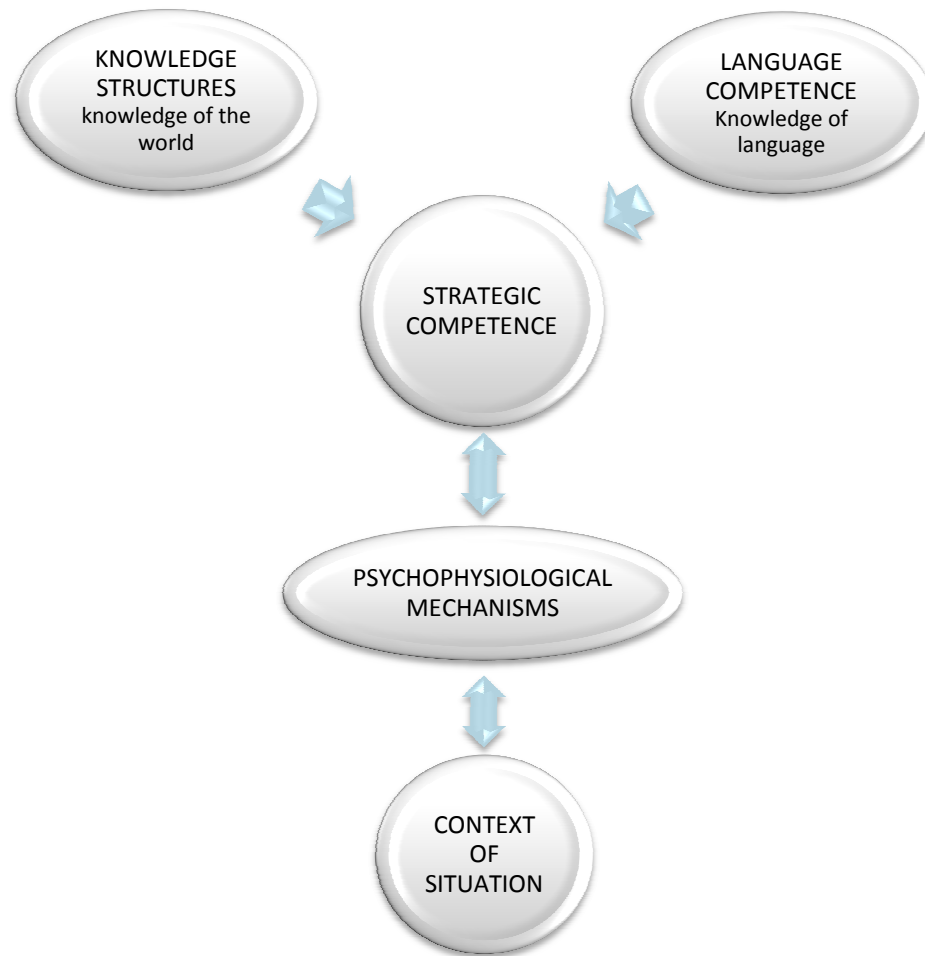


Figura n. 2.1 – Il modello di Bachman (Bachman, 1990:84)

La competenza linguistica, a sua volta, viene suddivisa in competenza organizzativa e competenza pragmatica (fig. n. 2.2). La “competenza organizzativa” comprende tutte quelle competenze coinvolte nel controllo della struttura formale della lingua al fine di produrre frasi grammaticalmente corrette e ben organizzate per l’elaborazione di un testo. Si distinguono al suo interno la “competenza grammaticale” (a sua volta distinta in lessico, morfologia, sintassi, fonologia/grafematica) e la “competenza testuale” (a sua volta distinta in coesione e organizzazione retorica del discorso). La “competenza pragmatica” riguarda la conoscenza delle convenzioni da rispettare nella produzione di un messaggio che sia appropriato in un dato contesto, e si distingue in “competenza

illocutoria” (conoscenza del rapporto tra un enunciato e l’atto o la funzione che si intende eseguire attraverso la produzione dell’enunciato stesso) e “competenza sociolinguistica” (capacità di eseguire le funzioni linguistiche in modo appropriato rispetto al contesto situazionale).

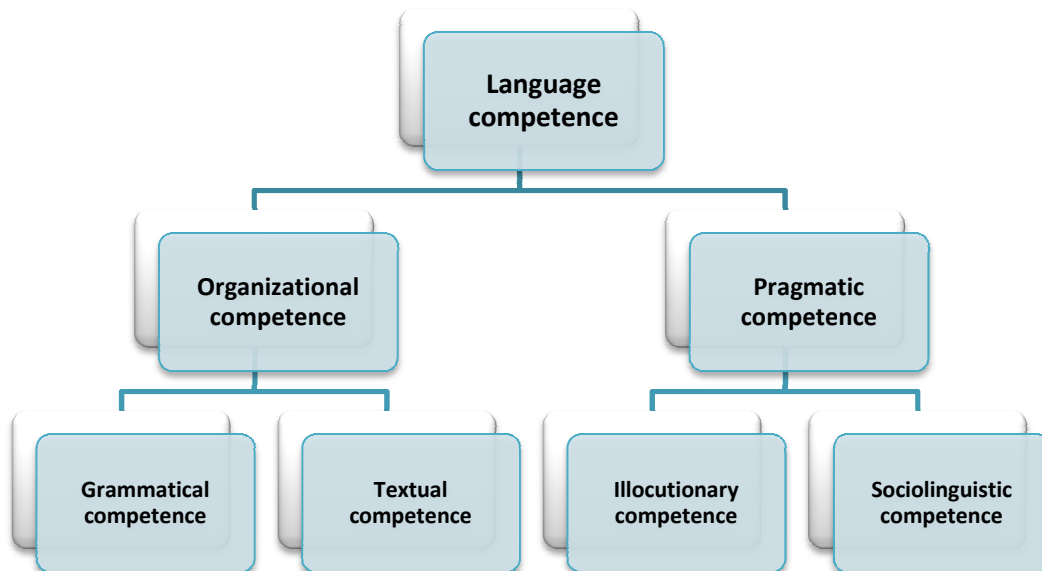


Figura n. 2.2 - Componenti della competenza linguistica

La “competenza strategica” è una competenza a sé stante (diversamente da Canale e Swain che la consideravano come un aspetto della conoscenza linguistica), che permette di pianificare ed eseguire un’interazione linguistica appropriata al contesto in cui ha luogo, utilizzando i mezzi più efficaci al fine di raggiungere lo scopo prefissato. Essa comprende le seguenti componenti: valutativa (affinché si ottenga una conversazione è necessario non solo determinare la propria conoscenza della lingua ma anche quella del proprio interlocutore, e accertarsi di condividere la conoscenza sull’argomento del discorso), di pianificazione (formulare un piano, attraverso il recupero degli elementi della competenza linguistica, la cui realizzazione permette il raggiungimento dello scopo comunicativo) e di esecuzione (mettere in pratica il

piano scegliendo modalità e canali appropriati al contesto).

La componente esecutiva ricorre ai “meccanismi psicofisiologici”, che si possono distinguere in canale visivo e uditivo, e modo produttivo e ricettivo. Bachman sostiene che nell’uso ricettivo della lingua vengono impiegate le abilità uditive e visive, nell’uso produttivo le abilità neuromuscolari, ossia relative ai nervi e ai muscoli (ad es. i movimenti articolatori dell’apparato fonatorio).

La “*knowledge structures*” (fig. n. 2.1) è spesso riferita alla conoscenza del mondo, ed è una componente necessaria poiché fornisce informazioni che permettono di usare la lingua facendo riferimento al mondo in cui si vive.

Tutti i fattori sopra descritti influiscono nel processo comunicativo:

“In general, language use can be defined as the creation or interpretation of intended meanings in discourse by an individual, or as the dynamic and interactive negotiation of intended meanings between two or more individuals in a particular situation. [...] This discourse derives meaning not only from utterances or texts themselves, but, more importantly, from the ways in which utterances and texts relate to the characteristics of a particular language use situation.

Language use involves complex and multiple interactions among the various individual characteristics of language users, on the one hand, and between these characteristics and the characteristics of the language use or testing situation, on the other.” (Bachman e Palmer, 1996:62)

e dunque influiscono anche in una *performance* durante una prova linguistica.

Qualche anno dopo Bachman, insieme con Palmer (Bachman e Palmer, 1996), rielabora il modello appena descritto (fig. n. 2.1), e lo integra con fattori nuovi, rendendone esplicite le interazioni reciproche.

“Because of the complexity of these interactions, we believe that language ability must be considered within an interactional framework of language use. [...] focuses on the interactions among areas of language ability (language knowledge and strategic competence, or metacognitive strategies), topical knowledge, and affective schemata, on the one hand,

and how these interact with the characteristics of the language use situation, or test task, on the other.” (Bachman e Palmer, 1996:62)

La rappresentazione grafica del modello (fig. n. 2.1) viene riproposta come segue:

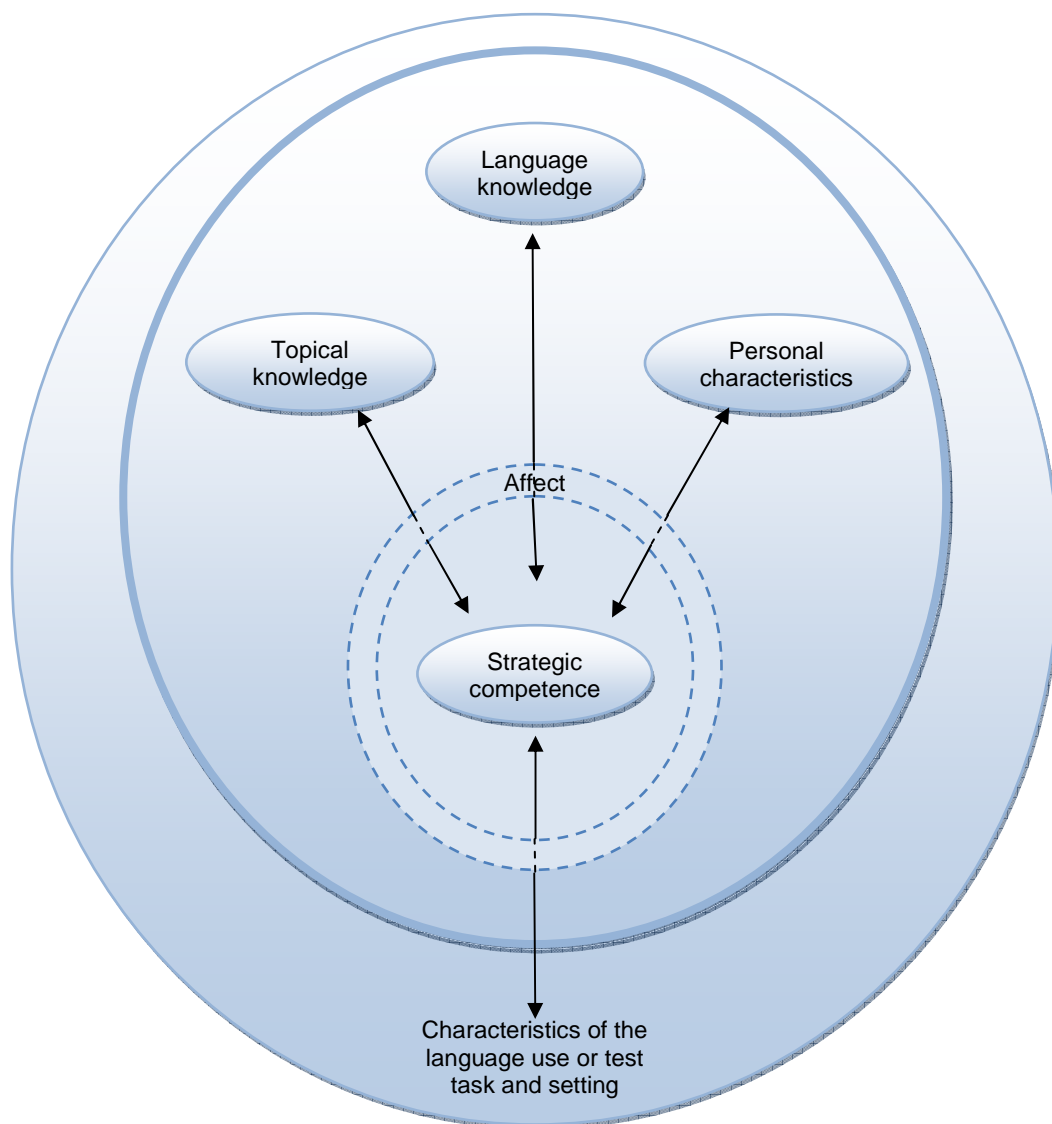


Figura n. 2.3 - Componenti dell'uso della lingua e della performance (Bachman e Palmer, 1996:63)

I fattori innovativi sono costituiti dalle caratteristiche personali e dagli schemi affettivi. La *knowledge structures* viene rinominata come “*topical knowledge*” e il *context of situation* diventa “*characteristics of the language use situation or test task*”.

Le “caratteristiche personali” comprendono quegli attributi dell'individuo (età, sesso, nazionalità, lingua materna, educazione, ecc.) che non fanno parte della competenza linguistica, ma che possono influenzare una *performance*.

Gli “schemi affettivi” costituiscono il nodo centrale tramite cui l'individuo valuta, consciamente o inconsciamente, le caratteristiche del compito e il contesto, in base alle proprie esperienze emotive vissute in altri contesti simili.

Questa rappresentazione, precisano Bachman e Palmer, costituisce un metodo per esprimere visivamente alcune delle principali interazioni (indicate dalle frecce a due punte) che hanno luogo nell'uso della lingua. Gli elementi situati nel cerchio più piccolo (evidenziato dal bordo più spesso) rappresentano le caratteristiche individuali di coloro che sono coinvolti nella comunicazione. Il cerchio esterno (più grande) comprende invece le caratteristiche del compito o della situazione con le quali l'individuo interagisce durante l'uso della lingua. La competenza strategica è posta al centro come anello di congiunzione tra le caratteristiche interne all'individuo, e tra esse e le caratteristiche del compito o della situazione.

Tale modello è concepito dagli autori come base concettuale e teorica per l'organizzazione sistematica delle componenti che intervengono in un processo comunicativo, tale da consentire una definizione dell'area di competenza da testare, utile per lo sviluppo pratico delle prove e per l'analisi delle abilità e dei fattori che agiscono in una *performance* durante sia la somministrazione che la valutazione del test.

Lo scopo di ogni attività di valutazione linguistica è quello di testare una competenza nascosta o difficilmente osservabile, come nella fattispecie la competenza comunicativa.

Le modalità per testare tale competenza consistono nell'osservazione del comportamento dell'esaminato durante una situazione valutativa e nella deduzione della conoscenza linguistica posseduta. Le inferenze generalmente si basano sull'interpretazione dei risultati di un test.

Mediante il modello di Bachman è possibile valutare tutte le componenti coinvolte durante una comunicazione tra due interlocutori e verificare la competenza linguistica di un candidato, determinando se la *performance* ha una corrispondenza con l'uso della lingua in una situazione di vita reale, piuttosto che essere fine a se stessa (ossia al superamento del mero esame).

Durante la fase valutativa è necessario tenere in considerazione il fatto che la *performance* è un esempio specifico dell'uso della lingua, che il candidato è un individuo parlante in un contesto valutativo, e che la prova linguistica è esattamente corrispondente all'uso della lingua in una specifica situazione linguistica.

Bachman e Palmer illustrano queste corrispondenze attraverso lo schema seguente:

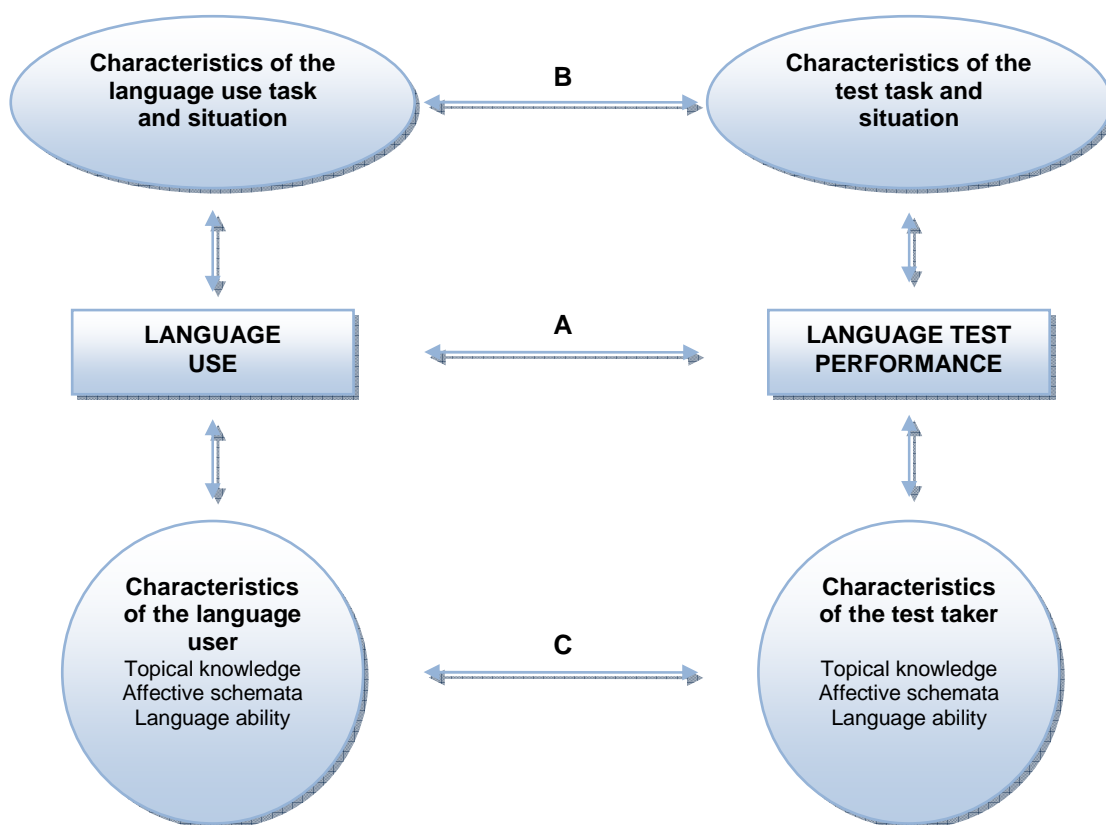


Figura n. 2.4 - Corrispondenze tra l'uso della lingua e la *performance* del candidato (Bachman e Palmer, 1996:12)

La corrispondenza centrale è quella stabilita tra l'uso della lingua in una *performance* e l'uso della lingua in una situazione non valutativa, come illustrato dalla freccia A in figura n. 2.4.

Le caratteristiche fondamentali che influiscono sull'uso della lingua, sia in una *performance* che in una situazione di vita reale, sono relative alle caratteristiche del compito, che deve essere attinente a una situazione valutativa oppure a una non valutativa (freccia B, fig. n. 2.4), e alle caratteristiche dell'individuo, sia che si tratti di un parlante di una comunità che di un candidato a un esame linguistico (freccia C, fig. n. 2.4).

La caratteristica più rilevante dell'individuo è la competenza linguistica, dal momento che costituisce ciò che si vuole misurare in una prova linguistica (le altre caratteristiche vanno considerate poiché possono facilitare la *performance*

del candidato).

Lo scopo del *testing* è di valutare la reale competenza linguistica, pertanto un test ben costruito, utile alla valutazione, implica la necessità di far corrispondere la *performance* con la competenza del candidato nell'uso della lingua al di fuori del contesto dell'esame.

2.4 L'utilità e le qualità essenziali di un test

La considerazione più importante nella progettazione e nello sviluppo di un test linguistico è l'uso cui è destinato, pertanto la qualità più importante di un test è l'**utilità**, il suo essere adatto allo scopo prestabilito. Ciò che rende utile un test è stato ed è ancora oggetto di studio e ricerca nel campo del *testing* linguistico.

La tradizione linguistica si è concentrata per lo più su due qualità di misurazione dell'utilità che sono la validità e l'affidabilità, presentate spesso in conflitto l'uno con l'altra, come se una qualità potesse essere rispettata solo a discapito dell'altra. In questo modo un test sarebbe il risultato di un compromesso tra validità e affidabilità.

Recentemente tra gli studiosi ed esperti di *testing* è maturata l'idea che la validità, l'affidabilità e altre qualità essenziali non possono essere valutate separatamente, bensì ognuna interagisce con le altre, ed è necessario raggiungere un equilibrio adeguato tra di esse nell'interesse dell'utilità complessiva del test.

Bachman e Palmer (1996) suggeriscono un modello come punto cardine per il controllo della qualità di ogni test, da considerare durante l'intero processo

di progettazione e sviluppo delle prove linguistiche. Tale modello di utilità comprende sei qualità essenziali (affidabilità, validità di costrutto, autenticità, interattività, impatto e praticabilità), tra loro complementari. L'utilità non può essere valutata in generale per tutti i tipi di test, né tanto meno è possibile definire il giusto equilibrio tra queste qualità in modo universale. Per ottenere un test utile è necessario ricercare un equilibrio appropriato tra le qualità, che varia per ogni specifica situazione valutativa. L'equilibrio è determinato dalla combinazione delle differenti qualità, che incidono sull'utilità complessiva di un determinato test. L'utilità di un test è stabilita in maniera soggettiva dal valutatore, poiché richiede un giudizio personale nella scelta dell'equilibrio delle qualità, di cui nessuna va ignorata. Pertanto, per lo sviluppo di ogni singolo test, si deve avere in mente lo scopo specifico, la particolare tipologia di candidati e lo specifico dominio d'uso della lingua (Bachman e Palmer si riferiscono a tale dominio attraverso l'espressione "*target langue use -TLU- domain*", e al compito in un dato dominio con "*TLU task*").

Delle sei qualità della rappresentazione di Bachman e Palmer:

Usefulness: Reliability + Construct validity + Authenticity + Interactiveness + Impact + Practicality

Figura n. 2.5 - Utilità (Bachman e Palmer, 1996:18)

le prime due sono propriamente qualità di misurazione, attraverso le quali si possono fare inferenze sulle competenze del candidato, poiché forniscono un riferimento per l'interpretazione dei risultati; contribuiscono a ridurre al minimo gli effetti dell'errore di misurazione se debitamente prese in considerazione; contribuiscono a portare al massimo l'effetto delle capacità linguistiche che un test si prefigge di misurare. La valutazione dell'utilità di un test dipende considerevolmente dalla *performance* del candidato, in particolare dai processi e dalle strategie che egli attiva per rispondere ai compiti di un dato test, e dall'esito

di tali processi, ossia dai risultati ottenuti dallo svolgimento del test.

2.4.1 Affidabilità

L'affidabilità (*reliability*) indica la coerenza nella misurazione dei risultati. Un test affidabile riproduce risultati costanti in somministrazioni successive per lo stesso candidato. Se un test è affidabile, somministrato allo stesso gruppo di individui in due momenti diversi e in due contesti differenti, produce gli stessi risultati.

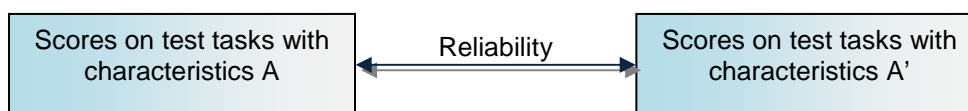


Figura n. 2.6 – Affidabilità (Bachman e Palmer, 1996:20)

Nella figura n. 2.6 la freccia a due punte indica la corrispondenza tra le due somministrazioni (A e A') in termini di punteggio. Questo metodo è detto "test-retest", poiché prevede la somministrazione dello stesso test a uno stesso gruppo di candidati in due momenti successivi, previa verifica delle condizioni nelle quali operano i candidati, che devono rimanere eque e non vi deve essere il rischio che intervengano elementi che possano disturbare o alterare i risultati.

Anche due forme equivalenti dello stesso test (*parallel form*), se create per essere utilizzate interscambiabilmente, per essere definite affidabili, devono poter produrre risultati uniformi, se somministrate allo stesso individuo contemporaneamente. Questo secondo metodo presenta però la difficoltà di garantire l'effettiva equivalenza tra le due forme del test.

Per porre rimedio a questa difficoltà si può fare ricorso al metodo del "semitest" (cosiddetto metodo "split-half"), ossia alla somministrazione delle due metà di un unico test (ad es. disponendo i quesiti pari da una parte e i quesiti

dispari dall'altra), come se fossero due test paralleli³. Questo metodo presenta, rispetto ai primi due, un notevole vantaggio: prevede una sola somministrazione. Al contrario i primi due necessitano di due somministrazioni, peculiarità che può causare alcuni svantaggi in termini di calo di interesse e motivazione da parte del candidato nel dover ripetere una seconda volta il test, o in termini di condizioni ambientali sfavorevoli o dissimili dalla precedente somministrazione. Tuttavia il metodo *split-half* presenta un inconveniente nel momento in cui il test non si presta alla divisione in due parti uguali per via della dipendenza di un *item* dall'altro. In casi come questi i primi due metodi risultano più appropriati ed efficaci.

Il conseguimento, da parte di un individuo, di risultati diversi in due test, somministrati attraverso uno dei tre metodi sopra descritti, non permette di giudicare tali risultati come indicatori affidabili del grado di competenza posseduta dall'individuo. Non è possibile stabilire quale dei due esiti è il reale riflesso della competenza, dunque entrambi gli esiti sono da considerare inaffidabili. Al contrario, l'esito positivo della correlazione tra le due somministrazioni può confermare l'effettiva coerenza interna del test.

È chiaro che non è possibile eliminare completamente la variabilità dei risultati, che può dipendere da fattori che vanno al di là del test e che possono afferire al candidato e alle condizioni in cui avviene il test. Per ridurre al minimo la variabilità dei risultati si può tentare l'identificazione delle potenziali fonti di errore; la diminuzione degli effetti delle diverse fonti di incoerenza già durante la fase di progettazione del test; infine la diminuzione degli errori di valutazione durante la fase di analisi dei risultati.

³ Per maggiori chiarimenti sui tre metodi di valutazione dell'affidabilità si rimanda a Messick e Douglas, 1967:226-235.

2.4.2 Validità di costrutto

La validità (*validity*) indica la condizione di un test di essere in grado di misurare ciò che intende misurare.

Bachman (1990) sostiene che la validità è una qualità relativa all'attendibilità e all'appropriatezza di interpretazione dei risultati di un test, a differenza dell'affidabilità che si riferisce ai punteggi stessi del test, alla misurazione dei risultati.

Bachman (1990) scompone la validità in varie componenti (saranno analizzate in seguito nel paragrafo 2.8.1) e focalizza l'attenzione (1996) sulla validità di costrutto (*construct validity*) per l'elaborazione del modello di utilità (illustrato in fig. n. 2.5).

Per dimostrare che un determinato punteggio o giudizio rifletta l'ambito di capacità linguistica che ci si prefigge di misurare è possibile definire un costrutto⁴ di ciò che si vuole misurare. Un particolare costrutto può essere considerato un valido indicatore di misura della competenza linguistica, esso fornisce la base di riferimento per un dato test e per l'interpretazione dei risultati ottenuti da questo test. Se i risultati ottenuti sono in linea con le previsioni fatte in precedenza il test è valido.

La validità di costrutto concerne anche il dominio di generalizzazione, al quale si estendono, procedendo oltre i dati osservabili, l'interpretazione personale dei risultati e le inferenze relative alla capacità possedute dagli individui. Infatti l'interpretazione sulla capacità linguistica si estende sempre oltre

⁴ Il costrutto può essere visto come la definizione di ciò che si vuol misurare, la definizione di una competenza o di un aspetto della competenza linguistica, in cui si specifica come essa sia collegata alle *performance* osservabili. Permette di ipotizzare le relazioni che possono esistere tra la competenza linguistica e il comportamento osservato (Cronbach e Meehl, 1955:283).

la situazione valutativa di un particolare dominio.

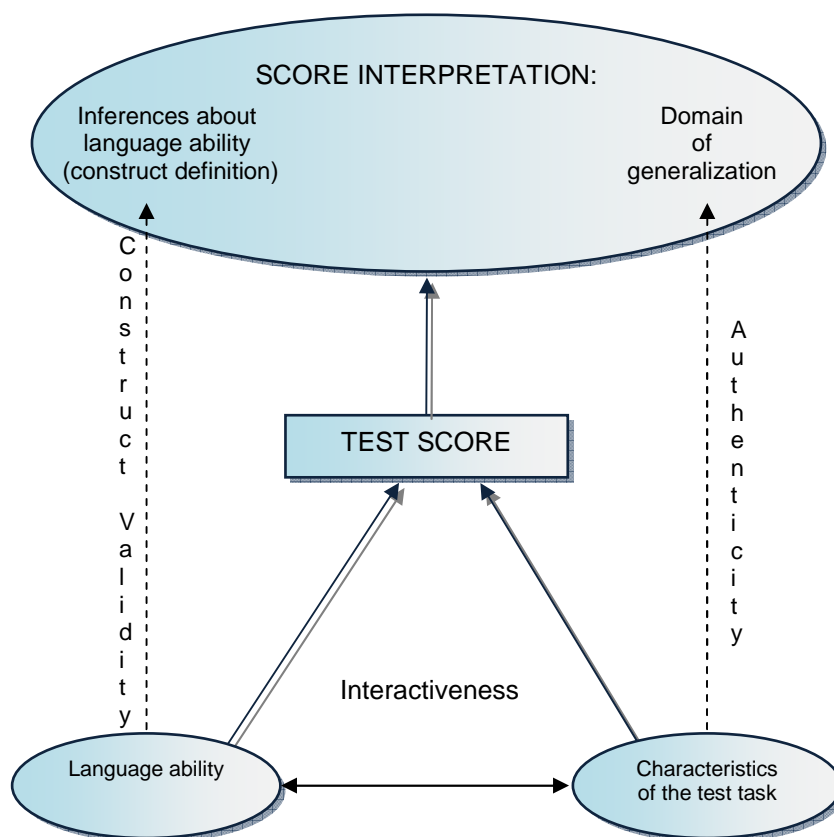


Figura n. 2.7 – Validità di costrutto (Bachman e Palmer, 1996:22)

La figura n. 2.7 illustra gli aspetti che caratterizzano l'interpretazione dei risultati: la validità di costrutto, l'autenticità e l'interattività. Quando si considera la validità di un test bisogna valutare sia il costrutto, sia le corrispondenze tra le caratteristiche del compito e quelle del dominio di generalizzazione (paragrafo 2.4.3 per il concetto di autenticità), sia il grado di interazione tra il compito e le caratteristiche individuali del candidato (paragrafo 2.4.4 per il concetto di interazione).

Per questo motivo è evidente che i risultati di un test non possono essere sempre considerati validi in modo assoluto. Ogni test va valutato in modo relativo sulla base delle componenti coinvolte di volta in volta nell'intero processo illustrato nella figura n. 2.7.

Inoltre va ricordato che attraverso le verifiche empiriche e le relazioni tra i costrutti e i risultati del test si può solo inferire sul grado di competenza linguistica. Ciò a volte può risultare riduttivo per due ragioni. In primo luogo un costrutto non sempre comprende l'intera area della competenza linguistica che si vuole misurare ma può essere limitato a un solo aspetto di essa. In secondo luogo è possibile, attraverso le verifiche e le interpretazioni dei risultati, dedurre la competenza di un individuo in modo astratto; fondamentalmente non si ha una prova evidente della competenza in tutti i domini e su tutti i costrutti possibili.

2.4.3 Autenticità

L'autenticità (*authenticity*) è data dal grado di corrispondenza tra le caratteristiche di un compito e le caratteristiche di una situazione reale in cui si presuppone l'uso della lingua.



Figura n. 2.8 – Autenticità (Bachman e Palmer, 1996:23)

La relazione rappresentata nella figura n. 2.8 (che coincide con la corrispondenza della freccia *B* in fig. n. 2.4 tra l'uso della lingua e la *performance* del candidato) mette in correlazione le caratteristiche di un compito in uno specifico dominio d'uso reale della lingua e le caratteristiche di un compito in una specifica situazione valutativa.

L'autenticità di un test è importante perché può essere un mezzo per esaminare fino a che punto l'interpretazione dei risultati, e dunque la valutazione della competenza acquisita si estende al di là della *performance*.

Inoltre l'autenticità di un test è fondamentale per l'impatto che può avere

sul candidato e sulla sua *performance*. La similitudine tra una situazione valutativa e una situazione concreta e la spendibilità in contesti reali della competenza linguistica acquisita, forniscono al candidato un motivo per reagire positivamente al compito che gli viene somministrato, e di conseguenza per eseguire la *performance* al massimo delle proprie potenzialità.

Il rispetto dell'autenticità deve essere tenuto ben a mente durante la progettazione del test, e garantito durante l'intero processo di esecuzione del test da parte del candidato.

2.4.4 Interattività

L'interattività (*interactiveness*) si riferisce alla portata e al tipo di coinvolgimento delle caratteristiche individuali del candidato nello svolgimento di una prova.

Le caratteristiche dell'individuo che maggiormente vanno considerate sono la competenza linguistica, la conoscenza del mondo, le caratteristiche personali e gli schemi affettivi (fig. n. 2.3).

Il modo in cui queste caratteristiche partecipano durante lo svolgimento di una prova caratterizza l'interattività.

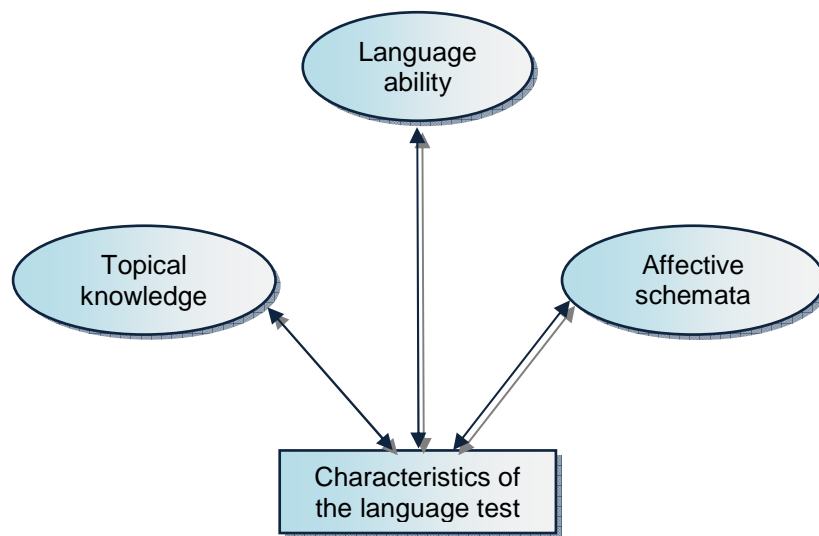


Figura n. 2.9 - Interattività (Bachman e Palmer, 1996:26)

Nell'illustrazione le frecce indicano le interazioni delle caratteristiche individuali con il compito. L'interattività risiede in queste interazioni, ed è una qualità che caratterizza qualunque compito, sia in una situazione valutativa che nella vita reale.

L'autenticità e l'interattività sono delle qualità relative, un test può essere più o meno interattivo e più o meno autentico. Un test interattivo o autentico per un gruppo di individui può non esserlo per un altro gruppo. Nella progettazione di un test si può preventivare il grado di autenticità e di interattività, ma non lo si può anticipare in assoluto.

Inoltre va considerato che un test, anche se possiede un basso grado di autenticità e di interattività, può essere considerato comunque relativamente utile, l'importante che si rispetti la validità e l'affidabilità. L'autenticità e l'interattività influiscono qualitativamente, più che quantitativamente.

2.4.5 Impatto

L'impatto (*impact*) si riferisce all'effetto che un test può avere sulla società e sul sistema educativo (macro livello), e sugli individui (micro livello) di questo sistema. Come afferma Bachman (1990), i test “*are intended to serve the needs of an educational system or of society at large*”.

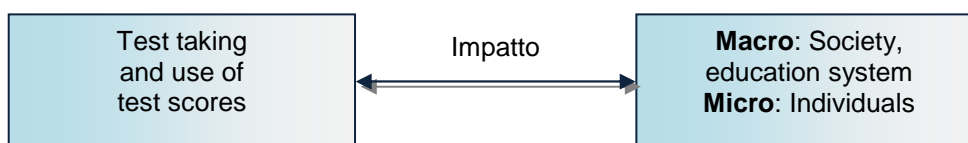


Figura n. 2.10 – Impatto (Bachman e Palmer, 1996:30)

I test sono progettati per raggiungere determinati scopi, i quali, una volta raggiunti, comportano delle conseguenze sia sugli individui coinvolti che sul sistema socioculturale.

In termini di micro livello l'effetto dell'impatto va valutato su una vasta varietà di individui, incide soprattutto in maniera diretta sugli *stakeholder* della formazione, ossia tutti coloro che sono esplicitamente coinvolti nel campo del *testing* (sviluppatori, amministratori, decisori politici, docenti e utenti finali). Qui preme sottolineare l'influenza che l'impatto ha sugli utenti finali, durante la fase sia di preparazione che di esecuzione del test, sia nel momento in cui si riceve il responso. Spesso infatti dall'esito di una prova dipende una decisione di grande importanza, ad esempio l'ammissione a un corso o a un'università, l'assunzione per un lavoro, ecc.

In termini di macro livello l'effetto dell'impatto va valutato in riferimento ai valori e agli scopi educativi di un sistema, che possono variare da una cultura a un'altra e nel corso del tempo.

Un importante aspetto dell'impatto è il "washback"⁵, ossia l'effetto potenziale che gli strumenti valutativi possono avere specificatamente sull'insegnamento e sull'apprendimento. È opinione condivisa, ad esempio, che la valutazione oggettiva basata sul quesito a scelta multipla può avere un washback negativo perché inibisce l'aspetto comunicativo della lingua, rispetto a una forma di valutazione come la *performance* di un compito realistico e concreto, che si basa sull'interazione e sull'uso comunicativo della lingua. Le forme di valutazione sono legate alla politica educativa e al sistema educativo nel complesso, in termini di metodi di insegnamento, di interazioni tra i compagni di una classe e tra essi e l'insegnante, di metodi di valutazione, e dunque di tutto ciò che riguarda utenti finali, docenti e sistema educativo di una determinata società.

Il carattere sociale della verifica linguistica è stato sottolineato negli ultimi anni, il test e le pratiche valutative non sono più considerate come attività isolate, ma come legate indissolubilmente alla politica educativa, pubblica e sociale. Pertanto i produttori di test e gli enti certificatori percepiscono, nello svolgimento delle loro attività, oggi più che mai, l'importanza delle loro scelte e la loro responsabilità etica e sociale.

2.4.6 Praticabilità

L'ultima delle sei qualità è la praticabilità (*practicality*), la quale possiede una natura diversa rispetto alle altre. Se le prime cinque qualità riguardano l'uso dei test, la praticabilità riguarda il modo in cui il test è realizzato e somministrato.

⁵ Un altro termine per riferirsi al washback è "backwash". Sembra però che nell'ambito del testing linguistico si preferisca il primo, dunque qui si utilizzerà sempre il termine "washback".

Se le risorse richieste per la preparazione di un test superano le richieste disponibili, il test è impraticabile, ossia impossibile da mettere in atto. Le risorse a cui si fa riferimento sono relative a materiali; tempi; costi; modalità di preparazione, somministrazione, correzione dei test; disponibilità di esperti e collaboratori.

Per poter determinare la praticabilità di un test bisogna considerare da una parte le risorse necessarie, e dall'altra l'assegnazione e la gestione delle risorse disponibili.

Bachman e Palmer (1996:36) definiscono la praticabilità come: "the relationship between the resources that will be required in the design, development, and use of the test and the resources that will be available for these activities", e ne rappresentano il rapporto in questi termini:

<p>Practicality: $\frac{\text{Available resources}}{\text{Required resources}}$</p> <p>If practicality ≥ 1, the test development and use is practical.</p> <p>If practicality < 1, the test development and use is not practical.</p>

Figura n. 2.11 – Praticabilità (Bachman e Palmer, 1996:36)

Se la richiesta delle risorse non supera la quantità di risorse disponibili, il test è realizzabile e si può procedere al suo sviluppo e utilizzo. Se la richiesta delle risorse supera le risorse disponibili, il test non è realizzabile e si deve procedere a una riduzione delle risorse richieste, oppure a un aumento delle risorse disponibili, o ancora a una riassegnazione delle risorse per un utilizzo più efficiente. Un test non dovrebbe mai superare i limiti delle risorse esistenti.

Le risorse disponibili, così come le risorse necessarie, variano comunque da una situazione a un'altra, per cui non è possibile determinare la praticabilità in modo universale.

2.4.7 Osservazioni finali

In conclusione le sei qualità sono tutte necessarie e vanno prese in considerazione insieme per valutare l'utilità di un test in una specifica situazione valutativa.

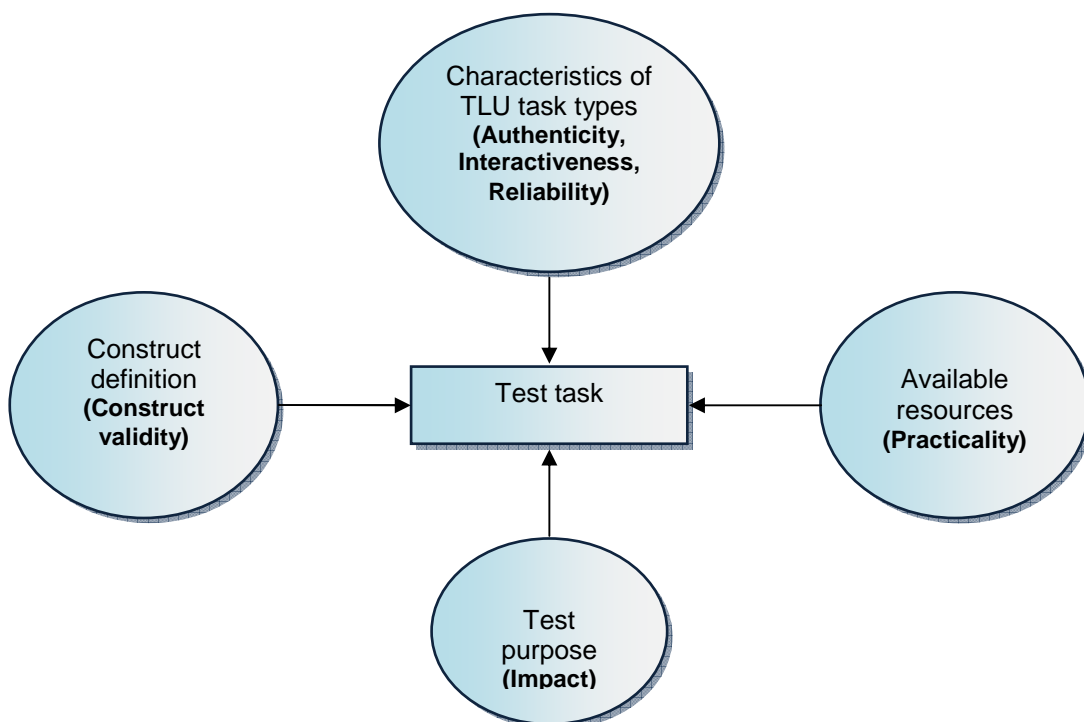


Figura n. 2.12 – Componenti e qualità dell'utilità (Bachman e Palmer, 1996:172)

Le qualità del test devono essere analizzate già durante le prime fasi di pianificazione e sviluppo, per minimizzare errori e modifiche successive.

Non è possibile, a ogni modo, raggiungere misure totalmente prive di errore (*error-free*) poiché ogni utilizzo del test è legato a fattori diversi, che non dipendono sempre e solo dalla costruzione del test stesso. È possibile invece raggiungere una misurazione il più possibile precisa (*more true score variance than error variance*⁶) attraverso la definizione di obiettivi e criteri di valutazione

⁶ Per maggiori approfondimenti sulla *True Score Theory* di Edgeworth e Spearman si rimanda a Wilson, 2005:83.

chiari e attentamente definiti. È possibile determinare un grado relativo di affidabilità, validità, autenticità, interattività, impatto e praticabilità, legato a un particolare contesto e che richiede, in ultimo grado, il giudizio dell'utente.

2.5 Il modello VRIP del Cambridge ESOL

L'Università di Cambridge *ESOL Examinations* provvede all'elaborazione e alla somministrazione di test e al rilascio di certificazioni di lingua inglese e di altre lingue su larga scala.

Essendo membro fondatore di ALTE (paragrafo 1.8), gli esami fanno riferimento al sistema internazionale di livelli di competenza utilizzato per la valutazione della conoscenza delle lingue europee; e il superamento degli stessi permettono di conseguire una certificazione che facilita il confronto con altre lingue, a vantaggio della mobilità educativa e professionale in Europa e oltre.

Nei primi anni Novanta il Cambridge ESOL ha sviluppato una prospettiva simile al modello di Bachman e Palmer sulla "utilità", basata su quattro qualità: validità, affidabilità, impatto e praticabilità (il cosiddetto "*VRIP model*"). Un test è valido se i risultati ottenuti sono il riflesso della capacità che il test si prefigge di misurare; è affidabile se i risultati ottenuti sono costanti e precisi, e quindi attendibili, ed è praticabile in base alle risorse necessarie per la sua realizzazione. L'impatto consiste nell'effetto (preferibilmente positivo), che ha il test sui candidati e sugli altri utenti, compresa la società in senso più ampio.

L'attività del Cambridge ESOL mira a garantire che i test rispondano a *standard* accettabili in relazione a queste quattro qualità essenziali. L'equilibrio delle qualità può differire in base allo scopo di ogni determinato test. In ogni caso

queste quattro qualità sono importanti in egual misura e il rispetto di esse dipende dalla definizione e dell'attuazione di specifici criteri.

I progettisti e gli sviluppatori di prove linguistiche hanno la responsabilità di fornire prova agli utenti finali della presenza delle qualità di affidabilità, utilità, ecc. e della correttezza delle interpretazioni dei risultati di un test. D'altra parte gli utenti finali hanno la responsabilità di esigere tali prove.

I progettisti, gli sviluppatori e gli utenti finali possono ricorrere a varie procedure per procurare e ottenere prove di utilità, sia qualitativamente attraverso approcci orientati allo studio sistematico dei fattori coinvolti nell'elaborazione di un test (ad es. l'*Evidence-Based Approach* di Weir descritto nel paragrafo 2.8.1), sia quantitativamente attraverso procedure statistiche (delle quali si approfondirà nel capitolo terzo).

2.6 Performance e uso della lingua: le caratteristiche del compito

Molte delle qualità essenziali di un test esaminate prima (paragrafo 2.4) derivano dalle corrispondenze tra la *performance* del test e l'uso della lingua in situazioni non valutative, discusse nel paragrafo 2.3 (fig. n. 2.4). Tali corrispondenze sono dimostrabili solo se è possibile descrivere e confrontare le caratteristiche salienti sia del compito richiesto dal test che del compito richiesto dall'uso reale della lingua.

Il grado di corrispondenza tra le caratteristiche del compito e quelle dell'uso della lingua in contesti concreti determina l'autenticità del compito, la validità delle inferenze compiute e il dominio a cui le inferenze si estendono.

Poiché la *performance* deve essere il riflesso della reale competenza del candidato e poiché le caratteristiche del compito⁷ fanno scaturire un certo comportamento in una *performance*, si rende necessaria una progettazione del compito che sia il più possibile simile a una situazione reale (tale da poterlo considerare un esempio specifico dell'uso concreto della lingua), e dunque si sottintende un'attenzione particolare verso tali caratteristiche nello sviluppo di un determinato compito.

Un quadro di analisi delle caratteristiche rilevanti del compito viene presentato da Bachman (il cosiddetto "*Framework of language task characteristics*" in Bachman e Palmer, 1996:47-57; e il cosiddetto "*Framework of test method facets*" in Bachman, 1990:116-152), in cui si distinguono principalmente le seguenti caratteristiche: *setting*; *rubric*; *input*; *expected response*; rapporto tra *input* e risposta attesa. Si tratta di caratteristiche, tra le tante che possono influenzare una *performance*, che stanno sotto il diretto controllo degli sviluppatori di un test, la cui comprensione e gestione danno vita a un test dalla massima utilità.

Il quadro dell'autore si presenta come uno strumento flessibile da adattare e modificare in base alle diverse situazioni e ai diversi scopi per cui un test è progettato.

Le caratteristiche del "*setting*" comprendono l'ambiente fisico (luogo, livello di rumore, temperatura, illuminazione, familiarità con i materiali e le apparecchiature a disposizione), i partecipanti (somministratori e candidati, familiarità tra essi) e il tempo che viene concesso ai candidati per lo svolgimento

⁷ Il compito, secondo Bachman e Palmer, è "*an activity that involves individuals in using language for the purpose of achieving a particular goal or objective in a particular situation.*" (Bachman e Palmer, 1996:44).

della prova.

Le caratteristiche delle “*rubrics*” includono la struttura del test (descrizioni relative al numero delle parti di un test, alla sequenza delle varie parti, al numero di *item* per ogni parte, alla diversa rilevanza delle varie parti), le istruzioni (lingua e canale tramite cui sono presentate, spiegazioni sulle procedure da seguire), la durata di ogni singolo compito del test e i metodi di valutazione e assegnazione del punteggio (criteri di correttezza, procedure di valutazione, quali ad es. il numero dei valutatori).

Le caratteristiche dell’“*input*”, ossia del materiale contenuto in un compito, riguardano il formato in cui l’*input* è presentato (il canale - visivo/uditivo-, la forma - linguistica/non linguistica-, la lingua - nativa/*target*-, la lunghezza, la tipologia - *item/prompt* - tramite cui si richiede una risposta limitata o estesa, il grado di velocità, il mezzo - dal vivo/riprodotto) e la lingua (caratteristiche organizzative e grammaticali, testuali; caratteristiche pragmatiche, funzionali e sociolinguistiche; caratteristiche dell’argomento).

Le caratteristiche della “*expected response*” (risposta attesa), ossia della risposta che si cerca di ottenere dal candidato (distinta dalla risposta effettiva, cosiddetta “*actual response*”), contemplano il formato (il canale - visivo/uditivo-, la forma - linguistica/non linguistica-, la lingua - nativa/*target*-, la lunghezza, la tipologia - di selezione/di produzione-, il grado di velocità, il mezzo - dal vivo/riprodotto-), e la lingua (uguali a quelle discusse per la lingua dell’*input*).

Le caratteristiche del “rapporto tra *input* e *expected response*” si riferiscono alla reattività (reciproco, non reciproco, adattivo⁸), allo scopo

⁸ I test di tipo adattivo (*adaptive*) non forniscono un *feedback* sulla correttezza delle risposte, tuttavia comportano una certa interazione poiché la domanda successiva posta al candidato dipende dalla risposta precedente. Per approfondimenti sulle differenze di reattività si rimanda a Bachman, 1990:148-152; Bachman e Palmer, 1996:54-57.

(ampio/ristretto, si riferisce alla quantità dell'*input* fornito) e all'immediatezza (diretto/indiretto) del rapporto.

Si ritiene che simili panoramiche di analisi delle caratteristiche del compito siano importanti perché forniscono modelli sulla base dei quali è consentita la descrizione e il confronto delle caratteristiche tra l'uso della lingua in una situazione reale e l'uso della lingua in una situazione valutativa, al fine di progettare nuove prove di valutazione e nuovi metodi valutativi sulla base della manipolazione delle caratteristiche evidenziate nell'analisi.

2.7 Le fasi di sviluppo del test

Le fasi di sviluppo di un test sono importanti in quanto, attraverso un attento disegno, si provvede alla creazione di un test che rispetti il principio dell'utilità per il raggiungimento dello scopo a cui il test stesso è destinato. Inoltre avere cognizione del processo da compiere consente di chiarire che cosa si fa, la ragione per cui si fa e il metodo che si adotta per farlo.

Lo sviluppo di un test è un processo che inizia con la programmazione e termina con l'utilizzo concreto del test. Bachman e Palmer (1996:87) delineano tre fasi principali dello sviluppo: la progettazione, l'operazionalizzazione (attraverso cui si attribuisce un contenuto empirico a concetti teorici astratti) e l'amministrazione del test.

Nella prima fase, durante la progettazione, innanzitutto si descrive (in un cosiddetto "*design statement*") lo scopo per cui è programmato il test, che nel caso di un test linguistico equivale in genere a fare inferenze sulla competenza linguistica e sulle abilità di un candidato. I passi successivi riguardano

l'identificazione e la descrizione dettagliata dei domini di contenuto da indagare, in modo da disegnare test autentici, i cui compiti corrispondono a usi specifici della lingua in situazioni reali (*TLU task*). Quindi si prosegue con la descrizione delle caratteristiche dei potenziali destinatari ai quali il test è destinato (caratteristiche personali, conoscenza del mondo, livello di educazione, ecc.). Successivamente si definiscono i costrutti che si vogliono misurare. Il valutatore può scegliere se includere, nella definizione del costrutto, in aggiunta alla definizione della competenza linguistica, anche le competenze strategiche e la *topical knowledge*. È indispensabile, inoltre, la compilazione di un piano per la valutazione dell'utilità. Esso è produttivo sia per il valutatore, durante tutte le attività che fanno parte di ogni fase del processo di sviluppo del test, sia per la definizione dell'equilibrio appropriato tra le sei qualità (paragrafo 2.4), al fine di ottenere un test utile (*"fitness for purpose"*). In ultima istanza è essenziale identificare le risorse necessarie e sviluppare un piano per la loro distribuzione e gestione (un esempio di tale piano si può trovare in Bachman e Palmer, 1996:161). Questo passo è importante perché rende evidenza della fattibilità del test.

Nella fase di operazionalizzazione, sulla base della trattazione sviluppata durante la progettazione, si sviluppano innanzitutto i compiti e le loro specificazioni (istruzioni, procedure di valutazione, ecc.). Dunque si comincia con lo sviluppo dei compiti sulla base dei *TLU task*. Tali compiti devono rispettare i principi di validità, affidabilità, autenticità, interattività, impatto e praticabilità, specificati nel piano per la valutazione dell'utilità, dunque i risultati prodotti da tali compiti dovranno essere il più possibile privi di errori di misurazione. Poi si sviluppa un progetto (cosiddetto *"blueprint"*) che descrive il modo in cui è costruito effettivamente ogni compito e il modo in cui ciascuno di questi è

organizzato per la realizzazione finale dell'intero test. Il secondo passo prevede la delimitazione delle istruzioni: fornire istruzioni chiare è di fondamentale importanza, poiché il candidato deve, attraverso la lettura di esse, comprendere la natura dell'intera procedura di verifica e di valutazione, l'abilità linguistica che si sta misurando, la struttura del test (il numero dei compiti di cui è composto, l'importanza di ognuno di essi, il tempo assegnato a disposizione per ciascuno di essi, ecc.), e i criteri adottati per la correzione dello stesso. Il candidato deve poter stabilire le procedure per lo svolgimento della prova per una resa ottimale della sua *performance*, e può farlo solo tramite una chiara comprensione delle istruzioni. Nella fase di operazionalizzazione si scelgono il canale e la lingua attraverso cui somministrare le istruzioni, e si opta per la possibilità o meno di fornire esempi di risposta dei compiti, per chiarire al candidato il modo in cui si deve rispondere ai vari compiti. Infine si sviluppano le procedure di correzione.

Bachman e Palmer (1996:193-226) forniscono due approcci per la valutazione. Il primo approccio prevede l'assegnazione di un punteggio per ogni compito completato con successo. In tal caso è necessario definire i criteri per stabilire ciò che costituisce il completamento di un compito con successo e determinare se ogni risposta è valutata come esatta/errata oppure vi sono vari gradi di correttezza. Il secondo approccio prevede la stesura di "*rating scale*", (scale di valutazione), nelle quali la competenza linguistica può essere valutata nel complesso, attraverso una serie di livelli che vanno dal livello zero (livello più basso, definito come "*no evidence of the ability*") al livello padronanza (livello più alto, definito come "*evidence of mastery of the ability*"); oppure scale in cui la capacità linguistica è distinta in varie componenti valutate separatamente. In entrambi gli approcci, i punteggi ottenuti dalle varie parti del test o dalle differenti scale analitiche, vengono uniti in un unico risultato composito (*composite score*). Questo punteggio può essere ottenuto tramite procedure di tipo compensativo

(*compensatory composite score*), ossia ottenuto dalla media o dalla somma di tutti i risultati, in cui i punteggi alti e quelli bassi si bilanciano e i punteggi alti compensano i punteggi bassi; oppure tramite procedure di tipo non compensativo (*non-compensatory composite score*), ossia il punteggio finale sarà uguale al punteggio più basso ottenuto nelle varie parti del test. Nel primo caso si presume che l'individuo possa, con livelli alti di competenza in alcune aree, compensare i livelli bassi di competenza in altre aree. Nel secondo caso l'individuo deve mostrare anche solo un livello minimo di competenza in ogni area, e i livelli alti di competenza in alcune aree non possono compensare i livelli bassi nelle altre.

Il momento di valutazione di un test può avere due scopi: la valutazione della validità del test in termini di contenuto e la valutazione della competenza del candidato. Il primo scopo deve precedere il secondo, e deve necessariamente essere raggiunto prima di giungere al secondo. Durante la fase di amministrazione del test si effettua un "*pre-test*" (verifica preliminare di collaudo) su un gruppo di individui, allo scopo di analizzare i risultati e verificare l'utilità e l'efficienza del test (si tratta di un "*mock test*", ossia di una simulazione d'esame). Lo sviluppo di un metodo di valutazione è un processo ciclico, la specificazione iniziale delle procedure e dei criteri, il *pre-test*, l'analisi dei *feedback*, l'eventuale revisione dei problemi emersi, la risomministrazione del test, la misurazione dei punteggi e così via. Solo attraverso la somministrazione-prova (*pre-test*) e l'analisi dei risultati ottenuti, è possibile verificare l'utilità e l'efficienza del test. Somministrare il test (sia per il *pre-test* che per il test reale) presuppone la preparazione dei materiali, la formazione degli esaminatori, la scelta del luogo, la gestione vera e propria della verifica e la raccolta qualitativa e quantitativa dei dati. Lo scopo del *pre-test* consiste nella raccolta di informazioni sull'utilità del

test e sulle procedure di somministrazione e gestione dello stesso, al fine di apportare le dovute modifiche. La raccolta di *feedback* è necessaria per stabilire l'appropriatezza dell'ambiente, della preparazione dei supervisori e del tempo assegnato, la chiarezza delle istruzioni e le reazioni dei candidati ai vari *item*. Il campione di candidati selezionato per il *mock test* deve rispecchiare il più possibile le caratteristiche e la preparazione dei candidati che realmente devono sostenere l'esame finale, e le condizioni in cui il *mock exam* ha luogo devono simulare le reali condizioni di somministrazione del test finale. È possibile analizzare qualitativamente la reazione dei candidati al test attraverso l'osservazione del loro comportamento durante l'esecuzione del test, oppure chiedendo loro, a fine prova, di definire le proprie percezioni personali riguardo il test, attraverso questionari, interviste, scale di valutazione, ecc. Statisticamente è possibile controllare in modo quantitativo la difficoltà del test o delle sue parti, la funzionalità dei compiti e l'affidabilità del test. La raccolta di *feedback* può riguardare anche gli amministratori e i supervisori che hanno preso parte alla somministrazione e gestione della verifica. Inoltre i candidati reali, utenti finali del test, possono fornire un *feedback* sull'utilità del test in base ai loro particolari bisogni. In seguito all'analisi dei *feedback*, la fase successiva prevede la revisione del test (qualora il risultato abbia rivelato la scarsa utilità del test) e, di conseguenza, la successiva ripetizione di un secondo *pre-test* sulla base delle modifiche effettuate. Al contrario, se i *feedback* rivelano una buona utilità del test, la fase successiva prevede la somministrazione reale del test. I compiti ritenuti validi possono essere archiviati al fine di poter creare una banca dati per lo sviluppo di test ulteriori. Una banca dati è un sistema che permette di conservare i compiti in maniera organizzata, accompagnati da descrizioni dettagliate sulla difficoltà degli *item*, sui livelli di appartenenza e sui tratti testati. I compiti da inserire nel *database* possono essere tarati su un'unica scala di misura al fine di

poterne facilitare il recupero per un utilizzo appropriato. Il secondo (autentico e principale) scopo del test si esplicita per la prima volta durante la somministrazione reale del test, ossia nel momento in cui si raccolgono informazioni al fine di poter inferire sulle competenze dei candidati. L'ultimo passo di questa fase è dunque la misurazione e la valutazione delle competenze dei candidati. Il tipo di misurazione dipende dalla natura della competenza che si vuole misurare e dal tipo di compito somministrato. È possibile fornire una valutazione attraverso l'assegnazione di numeri (metodo generalmente adottato per la correzione di prove oggettive, soprattutto per la verifica delle competenze di comprensione scritta e orale) oppure attraverso una stima qualitativa (metodo adottato normalmente per valutare le competenze di produzione scritta e orale).

Un simile quadro, che delinea le fasi dello sviluppo di un test, è stato elaborato dall'ALTE⁹ ed è attualmente utilizzato per l'elaborazione degli esami di certificazione delle lingue europee maggiormente studiate.

2.8 Fonti di errore nella valutazione

L'obiettivo principale di un test consiste nella misurazione delle competenze di un candidato. La misurazione dei risultati di un test avviene attraverso l'interpretazione dei risultati. Per una corretta interpretazione è necessario eliminare ogni possibile fonte di errore in grado di alterare la *performance* e dunque il risultato finale del test. Il primo obiettivo, per l'eliminazione delle fonti di errore, è l'identificazione delle potenziali fonti e la

⁹ Il documento, denominato "*Materials For The Guidance Of Test Item Writers*", è reperibile online in lingua inglese all'indirizzo <http://www.alte.org/downloads/index.php> (ultima consultazione: febbraio 2011).

riduzione, qualora possibile, del loro effetto sul test. Le fonti di errore possono afferire a fattori che si presentano regolarmente e sistematicamente, quali le caratteristiche del test stesso (che rimangono uniformi da una somministrazione a un'altra) oppure le caratteristiche dell'individuo, come ad esempio, lo stile cognitivo, il *background*, l'età, il sesso, ecc. (che variano da individuo a individuo ma sono perennemente presenti perché intrinseci all'individuo stesso in generale), e a fattori casuali non prevedibili, quali, ad esempio, cambiamenti relativi all'ambiente, allo stato emotivo del candidato, a irregolarità durante la somministrazione del test, ecc. Il risultato della variabilità di tutti questi fattori è l'esecuzione di una *performance* ogni volta diversa da parte di uno stesso individuo, e il conseguimento di un punteggio differente di volta in volta. Inoltre la *performance* di un individuo varia da quella di un altro perché i differenti fattori incidono in maniera diversa nei vari individui.

Durante la fase di progettazione del test uno degli scopi fondamentali, come è stato detto più volte, è quello di ridurre al minimo gli effetti delle caratteristiche del test, degli attributi personali e di tutti quei fattori transitori che non afferiscono direttamente alla competenza linguistica ma che incidono sulla *performance* e sulla produzione di risultati inaffidabili. L'analisi dei risultati del *pre-test* serve a verificare l'assenza di fonti d'errore, a controllare che le abilità vengano misurate in modo stabile e costante (affidabilità), ad appurare che ogni prova misuri le abilità che si prefigge di misurare (validità), e che ogni prova funzioni in un modo equo per tutti i candidati (assenza di *bias*).

Le caratteristiche personali possono essere, a volte, cause di "*bias*", ossia di discriminazione e pregiudizio verso fenomeni, ideologie e culture diverse rispetto alla propria. Spesso tali preconcetti interferiscono con la capacità di giudicare un determinato risultato o la *performance* di un determinato candidato

in modo imparziale e obiettivo, poiché si è influenzati da una prospettiva di tipo pregiudiziale. Se, ad esempio, alcuni candidati di una data nazionalità, messi a confronto con candidati di nazionalità diversa ma con la stessa abilità, risultano più o meno capaci in un dato test, la causa può essere attribuibile a *bias* operati nei confronti del gruppo di candidati appartenenti a una delle due nazionalità. Non si può parlare di *bias*, al contrario, se la differenza dei risultati è attribuibile a differenze o analogie linguistiche tra la lingua madre di uno dei gruppi di candidati e la lingua straniera testata. Se l'analisi quantitativa rivela possibili *bias*, è necessaria un'analisi qualitativa sugli *item*, i quali, se risultano affetti da *bias*, vanno rimossi o rettificati.

Le caratteristiche del test o i fattori transitori possono essere invece fonti di errore di misurazione, dunque da tener presente nella valutazione della validità e dell'affidabilità di un test. La constatazione di una significativa affidabilità rivela soltanto che lo strumento misura effettivamente qualcosa ma non fornisce alcuna indicazione sulla natura di ciò che si sta misurando. Dunque affrontiamo il processo di validazione esaminando più dettagliatamente le modalità di analisi e di verifica della validità e dell'affidabilità.

2.8.1 Validità e valutazione

La validità di un test esprime la capacità da parte di una procedura di misurazione di misurare ciò che si intende misurare. Un test può ritenersi valido se vi è una corrispondenza tra l'esecuzione del test e le inferenze fatte sulle capacità del candidato. Sebbene la validità sia un concetto unitario¹⁰, è possibile

¹⁰Cronbach (1980:99) sostiene che la validità è una. Ancora prima Messick e Douglas (1967:79) definiscono il concetto di validità come un concetto unitario sebbene sfaccettato.

distinguere al suo interno vari elementi che la compongono, i quali necessitano di considerazione nella progettazione e nell'uso di un test, e nell'interpretazione dei risultati che ne conseguono. Per stabilire se un test è valido, è necessario considerare una serie di aspetti relativi alla validità, che consentono di distinguere tra le categorie di validità di contenuto, di costruito, di criterio e validità esteriore (o di facciata).

Vi sono principalmente due modalità per verificare la validità di un test, il primo riguarda la validità di contenuto e di costruito e consiste nell'esame del contenuto del test; il secondo riguarda la validità di criterio e avviene statisticamente mediante la correlazione¹¹ con altre unità di misura.

La "validità di contenuto" (*content validity*) riguarda il grado di concordanza semantica tra il dominio di conoscenze e di abilità da misurare e gli indicatori scelti a rappresentarlo empiricamente. Tale validità si misura attraverso la verifica della qualità e del grado di rappresentatività degli indicatori che formano la misura. Gli indicatori dovrebbero mostrare una consistenza interna (ossia un accordo tra essi nel misurare la stessa abilità o lo stesso tratto di un'abilità), un confronto tra diverse applicazioni dello stesso strumento, e un rapporto di analogia tra la misura da validare e altri indicatori già validati. È necessario considerare le tabelle di specificazioni e gli *item* definiti nella fase di sviluppo del test, verificare in che misura gli *item* del test rappresentino l'universo degli *item* possibili e in che misura i compiti proposti rappresentino il costrutto misurato.

Perfino quando si è certi della validità di contenuto, non si ha la sicurezza che il test sia valido. È necessario un confronto con alcuni criteri che forniscono

¹¹ Per "correlazione" si intende un rapporto tra due misure, tale che al variare di una varia anche l'altra.

evidenza della validità.

La “validità di costrutto” (*construct validity*) riguarda il grado di precisione con cui un indicatore misura i tratti teoricamente postulati che si intendono misurare. Il costrutto costituisce la teoria soggiacente la progettazione del test, attraverso la somministrazione del quale si verifica l’acquisizione di tale costrutto. Per poter confermare la validità del test, dunque poter sviluppare procedure di misurazione appropriate per stimare l’avvenuta acquisizione della competenza definita dal costrutto o di un solo aspetto di tale competenza, è necessario tracciare, dal punto di vista concettuale, una definizione precisa dei costrutti, appropriati alle specifiche situazioni valutative. Dal punto di vista operativo il costrutto si definisce attraverso la scelta delle procedure di misurazione e di correzione dei risultati. La validità di costrutto può essere misurata attraverso l’analisi fattoriale e l’analisi di risposta degli *item*. L’analisi fattoriale è un metodo che permette di ridurre un alto numero di variabili ipotetiche a un numero minore di variabili osservabili, detti “fattori”. Nel caso di un test linguistico questa tecnica permette di ridurre il numero degli *item* identificando i fattori sottostanti condivisi da una serie di *item*. L’analisi di risposta degli *item* è un metodo per calcolare la probabilità che possiede un candidato di rispondere correttamente a ciascun *item*, in funzione del livello di competenza posseduto dal candidato stesso e dei parametri dell’*item* (L’analisi di risposta degli *item* si suddivide in varie tipologie, distinte per il numero dei parametri presi in considerazione. L’analisi più nota è basata sul modello di Rasch, detto anche “modello a un parametro”, in cui si prende in considerazione solo il livello di difficoltà dell’*item*, come si vedrà nel prossimo capitolo).

La validità di costrutto si distingue in “validità convergente” e “validità discriminante”. La prima è determinata dalla correlazione con i risultati ottenuti da

altre misure di costrutti teoricamente collegati con quello da misurare (ad es. prove inclini a testare una specifica abilità). La seconda tipologia di validità è data da una correlazione inconsistente con i risultati conseguiti con altre misure di costrutti diversi, non connessi con quello da esaminare (ad es. un test su una specifica abilità dovrebbe essere più simile a un altro che afferisce alla verifica della stessa abilità rispetto a un test che riguarda un'abilità diversa). La validità convergente e la validità discriminante si misurano mediante la “*multitrait-multimethod (MTMM) analysis*”¹² di Campbell e Fiske. I due studiosi sostengono che ogni costrutto (tratto) va misurato con metodi molteplici. La correlazione tra i risultati ottenuti dalla misurazione dello stesso tratto dovrebbe essere positiva, ossia i risultati dovrebbero convergere nonostante l'utilizzo di metodi diversi. Al contrario la correlazione tra i risultati ottenuti dalla misurazione di tratti diversi, anche se misurati con lo stesso metodo, dovrebbe essere nulla, ossia i risultati dovrebbero essere diversi. L'importanza del loro approccio sta nell'aver dato rilevanza, nella misurazione della validità, al metodo, oltre che al contenuto.

La “validità di criterio” (*criterion validity*) riguarda il grado di corrispondenza tra i risultati del test (del quale si intende assodare la validità) ottenuti attraverso un determinato indicatore e quelli ottenuti utilizzando un indicatore-criterio esterno (la cui validità è già stata dimostrata). Il miglior modo per stabilire la validità di criterio è la correlazione statistica tra i due risultati, ed è espressa dalla misura della varianza¹³ tra essi (la correlazione statistica e la varianza sono illustrati nel terzo capitolo).

La validità di criterio può essere concorrente o predittiva. Nella “validità

¹² Per approfondimenti maggiori si rimanda a Campbell e Fiske, 1959:81-105.

¹³ La “varianza” è una misura statistica che permette di capire quanto varia il risultato di un soggetto dalla media del gruppo.

concorrente” la correlazione tra l'indicatore da convalidare e l'indicatore-criterio avviene in concomitanza. Se i risultati risultano positivamente correlati, si possono ritenere validi. Nella “validità predittiva” l'indicatore da convalidare viene utilizzato in una fase precedente a quella in cui verrà utilizzato l'indicatore-criterio. Se i risultati convergono, rispetto ai risultati di verifiche successive, riconosciuti come criterio di validità, dunque risultano capaci di fare previsioni sulle competenze possedute, si possono ritenere validi.

La “validità esteriore” (*face validity*) riguarda il grado di percezione che gli *stakeholder* hanno sulla validità del test. La validità esteriore non sempre viene annoverata insieme alle tre maggiori categorie di validità precedentemente descritte. Mosier è stato il primo, nel 1947, a parlare di validità esteriore, descrivendone tre classificazioni. Il termine è caduto in oblio quando, nel 1985, l'*American Psychological Association* (1985b) ha riconosciuto la validità come distinta nelle quattro categorie di validità di contenuto, di concorrenza, di prevedibilità e di costruito. Bachman (1990:288) ha sostenuto la validità esteriore nel senso che ha riconosciuto l'importanza del modo di percepire un test da parte del candidato, per poter affrontare con serietà e motivazione il test stesso. Tuttavia è inappropriato, secondo lo studioso, servirsi della validità esteriore come metodo per approcciarsi al problema dell'utilità dei test.

2.8.2 Il modello di Weir

La tendenza più recente in merito al concetto di validità tenta di superare il puro modello meramente teorico, per orientarsi verso descrizioni più operative e pratiche di validazione. Tali descrizioni assumono la forma di quadri inclini a fornire precisazioni teoriche utili sia allo sviluppo di test che al controllo della

validità dei risultati. I quadri più rilevanti sono stati presentati da Mislevy nel 2003 (modello utilizzato come fondamento teorico per l'elaborazione degli esami TOEFL negli Stati Uniti) e da Weir nel 2005 (modello utilizzato per gli esami Cambridge ESOL in Europa).

Il modello di Weir, cosiddetto "*Test Validation Framework*" (2005a), risponde in buona parte alle stesse esigenze del modello sviluppato dal Cambridge ESOL (paragrafo 2.5), giacché si preoccupa del rispetto delle quattro qualità previste da quest'ultimo (validità, affidabilità, impatto e praticabilità).

La premessa alla base dell'approccio di Weir sta nel tentativo di fornire agli sviluppatori di un test una panoramica che possa favorire la ricerca di prove di validazione (*evidence*) sulla qualità del test, per rispondere al meglio alle richieste di utilità ed equità del test stesso (Weir propone una panoramica di riferimento del quadro, per ognuna delle quattro abilità, per illustrarne l'applicabilità in contesti operativi).

L'esigenza di raccolta di prove di validazione deve essere presente a ogni stadio dello sviluppo, della somministrazione e della valutazione del test.

L'intero quadro generale e le interazioni tra gli elementi che lo compongono sono illustrati nella figura seguente:

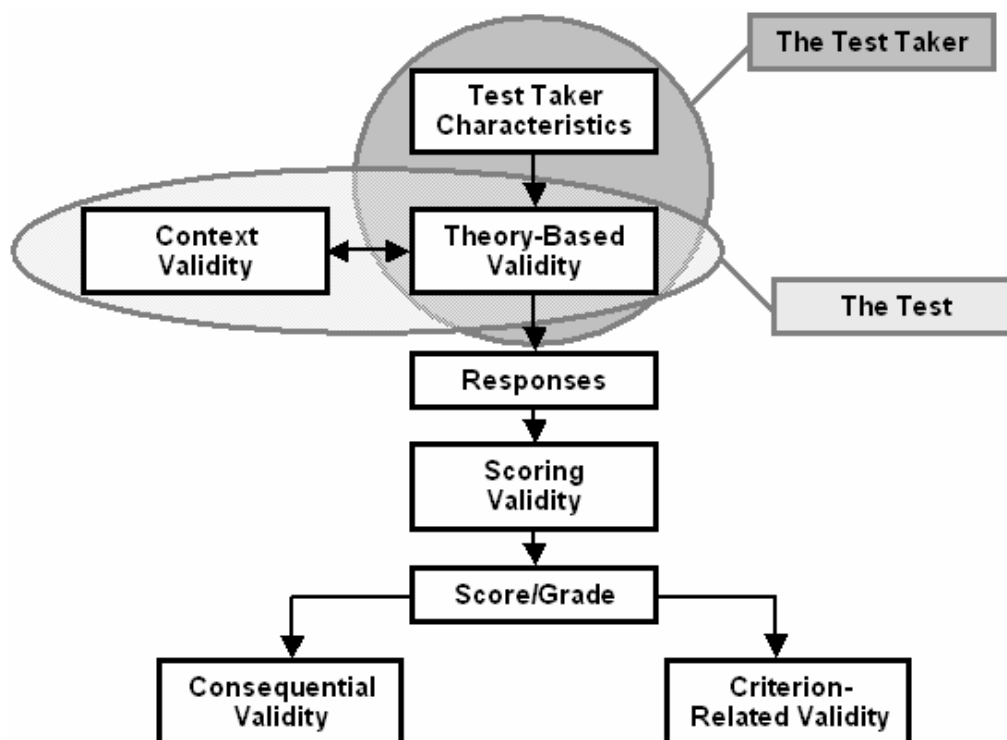


Figura n. 2.13 – Il modello di Weir (Weir:2005a)

Ogni elemento del quadro, come si può osservare in figura n. 2.13, è indipendente dall'altro. Tuttavia, Weir sostiene l'esistenza di un rapporto simbiotico tra la validità di contesto, la validità cognitiva e la validità di correzione, che a loro volta costituiscono nell'insieme la validità di costrutto. Dunque il concetto di validità viene riconfigurato e vengono palesate le interazioni che hanno luogo tra le sue sottocategorie. Ad esempio le decisioni prese in merito alle caratteristiche linguistiche e contestuali del compito hanno un impatto sul processo di elaborazione del compito stesso, e la comunicazione anticipata dei criteri di valutazione, ai candidati, influenza il loro modo di pianificare ed eseguire il compito.

Il quadro di riferimento proposto da Weir pone l'attenzione sulla dimensione cognitiva del candidato, sui processi mentali coinvolti nella dimostrazione delle competenze possedute o acquisite, e sull'aspetto sociale dell'uso della lingua, piuttosto che sul mero aspetto linguistico.

La rappresentazione grafica illustra come le varie componenti della validità combacino sia per quanto riguarda l'aspetto temporale che per quanto riguarda i concetti. Le frecce indicano le principali direzioni di ogni ipotetica relazione. Si parte dall'alto, dal momento in cui il test viene progettato, per essere poi somministrato, e infine corretto. È possibile prevedere in quale fase una componente della validità entra in gioco e che cosa comporta l'influenza di tale componente di validità ad ogni fase. Nelle fasi che precedono l'esecuzione del test entrano in gioco le componenti della validità di contesto e cognitiva, nelle fasi che seguono l'esecuzione del test entrano in gioco le componenti della validità di correzione, di criterio e consequenziale.

Weir sostiene che ogniqualvolta si progetta un test si devono considerare, per l'elaborazione di un test appropriato, le caratteristiche fisiche, psicologiche e esperienziali (ad es. età, *background* linguistico, educazione, stile d'apprendimento, motivazione, familiarità con il test, ecc.) del candidato (*test-taker characteristics*); i processi cognitivi richiesti per l'esecuzione del compito (*theory-based validity*); le caratteristiche del compito e le condizioni in cui il test è somministrato (*context validity*); le procedure e i criteri di valutazione (*scoring validity*); la correlazione con criteri esterni che possiedono una comprovata esperienza di misurazione (*criterion-related validity*). Infine vanno valutate le conseguenze che il test può avere sugli *stakeholder* e il grado di coerenza con gli scopi pedagogici e didattici del sistema educativo, e con i valori condivisi dalla società in cui il test ha luogo (*consequential validity*). La validità consequenziale è inerente ai concetti di *washback*, impatto e *bias* introdotti in precedenza.

Nell'approccio di Weir non appare il concetto di *content validity* (discusso prima e riconosciuto dall'APA come una delle componenti della validità), che si riferisce tradizionalmente al contenuto del compito, ossia all'universo di

contenuto misurato. In realtà Weir preferisce parlare di validità di contesto per indicare un concetto più esteso, che include la considerazione sia dei parametri di contenuto linguistico che dei contesti socio-culturali nei quali il compito viene eseguito.

La validità cognitiva è attinente all'affidabilità del test, poiché più si è capaci di descrivere in modo dettagliato il costrutto teorico che si vuol misurare (da qui l'espressione *theory-based validity* presente nella rappresentazione grafica) durante la progettazione del test, più efficienti saranno le procedure statistiche che verranno applicate ai risultati del test durante la fase di correzione. Lo stesso vale per la validità di correzione, la quale determina la coerenza e la stabilità dei risultati perché afferisce alla scelta di criteri adeguati e alla modalità di assegnazione dei punteggi condivisi tra i correttori.

I risultati di un test sono interpretati sulla base delle definizioni dei costrutti da misurare e dei domini specifici inclusi nel compito. Le inferenze sulla competenza linguistica vengono fatte sulla base dei costrutti e del dominio di generalizzazione a cui le inferenze si estendono. Le informazioni, che si ricavano dai risultati dei test, devono quindi riferirsi sia ai costrutti sia ai domini, e devono basarsi su criteri e scale di valutazione prestabilite.

Il metodo più adatto per verificare la validità di un test consiste nella correlazione tra le scale e i criteri di riferimento da una parte e i risultati del test dall'altra. Il coefficiente di correlazione tra i punteggi di riferimento e i punteggi grezzi ottenuti dal test è definito "coefficiente di validità" (il calcolo del coefficiente di correlazione è affrontato nel capitolo successivo).

2.8.3 Affidabilità e valutazione

I controlli di validità effettuati sui test sono privi di significato ed efficacia se non ci si accerta prima dell'affidabilità dei test stessi. Un test può essere valido, ossia misurare ciò che ci si era prefissati ma non produrre risultati costanti perché non ci si è preoccupati di limitare le possibili fonti di errore. Allo stesso modo un test può essere affidabile, ossia produrre risultati correlabili con altri test, ma non essere valido, ossia non misurare ciò che era stato stabilito. È pertanto inevitabile accertarsi necessariamente sia della validità che dell'affidabilità in modo interconnesso.

L'affidabilità di un test può dipendere da fattori interni o esterni al test. I fattori interni possono riguardare la chiarezza delle consegne, la qualità degli *item*, il numero e la loro discriminazione per livello di abilità, l'inadeguatezza del tempo concesso per lo svolgimento della prova. I fattori esterni possono riferirsi alle condizioni ambientali, alla personalità e professionalità dei somministratori, alle condizioni fisiche e psicologiche dei candidati. La variabilità di questi fattori interni ed esterni al test può pregiudicare lo svolgimento e dunque l'affidabilità del test. I fattori che compromettono l'affidabilità dei test possono riguardare anche, e soprattutto, la scarsa condivisione e l'utilizzo non sempre costante di misure di correzione appropriate e uniformi. Si può distinguere tra una "*inter-rater reliability*" quando vi è una coerenza di giudizio tra esaminatori diversi per uno stesso test, e una "*intra-rater reliability*" quando è presente una coerenza di giudizio da parte di uno stesso esaminatore in occasioni diverse e sotto circostanze diverse per uno stesso test.

Per verificare la presenza o l'assenza di fonti di errore che possono alterare la stabilità dei risultati si ricorre ai metodi descritti nel paragrafo 2.4.1 e successivamente si ricorre a strumenti statistici per calcolare la variabilità dei

risultati (di cui si approfondirà nel capitolo terzo).

2.9 Tipologie di *item* e scale di valutazione

Le verifiche e le pratiche di valutazione si modellano inevitabilmente al contesto storico e sociale in cui sono sperimentate e impiegate.

Dagli inizi del ventesimo secolo, negli Stati Uniti hanno dominato, nel campo del *testing* linguistico, i principi e le pratiche psicometriche, che a loro volta hanno influito sul *testing*, dal punto di vista della teoria e della pratica, in molte altre parti del mondo. La posizione dominante è stata assunta dalle prove di tipo obiettivo e dalle proprietà psicometriche dello strumento valutativo.

La tradizione britannica al contrario si è preoccupata più che altro delle questioni della validità, dal rapporto con il programma di studio e il curriculum studiorum dell'individuo, e delle relazioni tra pedagogia e apprendimento. Questo si riflette in una maggiore enfasi su un approccio basato sul compito e sulle abilità (piuttosto che sulle proprietà psicometriche), mediante prove atte a verificare le abilità di produzione scritta e produzione orale, le quali tendono a una valutazione più soggettiva.

La variabilità dei risultati risulta pertanto meno evidente quando si tratta di prove strutturate oggettive, ossia di verifiche costituite da domande chiuse che presuppongono delle risposte controllate. Esempi sono costituiti da prove di vario tipo come domande a scelta multipla (una serie di domande, ciascuna delle quali è dotata di più risposte, di cui una è corretta e le altre inesatte hanno la funzione di distrattori), vero/falso (una serie di domande, ciascuna delle quali prevede la scelta tra vero e falso), di completamento (una serie di frasi, ciascuna delle quali

presenta degli spazi vuoti in cui inserire una o più parole), di abbinamento (una serie di frasi da abbinare a frasi, definizioni o parole collegate alle prime).

La preparazione di prove di questo tipo richiede l'elaborazione di un certo numero di quesiti (*item*) che prevedono generalmente una sola risposta esatta (*key*) o un numero limitato e predefinito di risposte possibili (*mark scheme*).

Quanto più una prova è strutturata tanto più è attendibile, in quanto limita la possibilità che si verifichino errori di misurazione dovuti a correttori diversi o allo stesso correttore a distanza di tempo o in condizioni diverse.

L'oggettività di una prova è in ultima analisi data dal fatto che il punteggio per ogni risposta esatta è stabilito a priori; che nella misurazione non interferisce la soggettività del correttore (tanto che anche un computer può calcolare il punteggio conseguito da ciascun candidato); che tutti i candidati si trovano nella medesima condizione durante la somministrazione del test.

Le prove oggettive possono essere di tipo normativo (*norm-referenced test*) e di tipo criterio (*criterion-referenced test*). I test di tipo normativo prevedono l'interpretazione dei risultati sulla base di un confronto con i risultati ottenuti da un gruppo di individui (definito "*norm-group*"¹⁴) simile al gruppo a cui il test è stato somministrato. I test di tipo criterio prevedono l'interpretazione dei risultati sulla base di livelli di competenza o domini di contenuto presi come criteri di riferimento.

La correzione di una prova oggettiva è un'operazione relativamente rapida, si tratta di mettere insieme le risposte corrette, attraverso il conteggio di esse. Si procede infine all'attribuzione del punteggio grezzo totale. Questo

¹⁴ Il gruppo di individui è definito "*norm*" perché diviene lo standard (norma) per il confronto con i risultati di somministrazioni successive.

punteggio si ottiene moltiplicando il numero delle risposte esatte per ogni gruppo di *item* dello stesso tipo con il punteggio assegnato a priori per ogni *item*, e sommando dopo tutti i punteggi parziali.

Per ridurre le possibilità che il candidato possa rispondere a una domanda cercando di indovinarne quella corretta, piuttosto che rispondere perché effettivamente si conosce l'argomento (negli *item* vero/falso ad es. la possibilità di indovinare è pari al 50%), l'attribuzione del punteggio parziale per ogni singola risposta può essere effettuata attraverso la penalizzazione delle risposte errate. Si attribuisce un peso differente alle diverse risposte in base alla difficoltà della domanda o alla probabilità di rispondere correttamente, e si penalizzano gli errori (distrattori) attraverso la detrazione di una frazione di punteggio per ogni risposta errata. Poiché, invece, non viene applicata alcuna penalizzazione per le risposte omesse, il candidato si sente incoraggiato a non rispondere ai quesiti di cui non ha certezza della risposta esatta. Il punteggio con penalizzazione si calcola attraverso la formula della cosiddetta "ipotesi nulla":

Formula n. 2.1:
$$P = \left(E - \frac{S}{n-1} \right) K$$

in cui P il è il punteggio complessivo, E è il numero delle risposte esatte, S è il numero delle risposte sbagliate, diviso per n che è il numero delle alternative diminuito di 1, infine moltiplicato per K , ossia il peso attribuito a ciascun tipo di *item*.

Il punteggio grezzo, ottenuto dalla computazione delle risposte esatte (con o senza penalizzazione) necessita generalmente di una trasformazione in un giudizio che può essere interpretato in modo più efficace e significativo dal candidato, da un'istituzione o da chiunque è interessato a capire il grado di competenza del soggetto esaminato. Tendenzialmente i giudizi si esprimono

mediante valutazioni di tipo idoneo/non idoneo, promosso/respinto, giudizi che vanno dall'ottimo all'insufficiente, oppure attraverso uno schema convenzionale di lettere o numeri.

Per verificare se i punteggi grezzi di una prova sono affidabili, è necessario testare nuovamente gli stessi candidati somministrando loro una seconda prova, mediante uno dei tre metodi descritti precedentemente (*test-retest*, *parallel form* o *split-half*). Non si deve ottenere necessariamente lo stesso punteggio grezzo. È possibile stabilire punteggi di taglio e fasce di punteggi, in modo tale da poter collocare il risultato ottenuto nella stessa fascia di punteggi o avere un punteggio di taglio che cada nello stesso intervallo in entrambe le prove. Una volta stabilita la corrispondenza tra il primo e il secondo test, è possibile stabilire quale punteggio ottenuto in uno dei due test equivale al punteggio ottenuto nell'altro.

Le prove di tipo soggettivo tendono a produrre risultati meno affidabili rispetto alle prove oggettive. Si tratta di prove aperte in cui la valutazione è soggettiva (il più delle volte si tratta di prove per la verifica della produzione scritta e orale). Per rendere tali prove più eque, è necessario usare alcune accortezze durante la preparazione e la valutazione del test.

Innanzitutto la *performance* del candidato dipende dall'argomento scelto per la prova e dal tipo di compito. Preparare una vasta gamma di argomenti e una varietà di compiti fornisce al candidato un'ampia possibilità di scelta, da compiere sulla base delle proprie personali inclinazioni. Secondariamente (ma non per importanza) è necessario basare la valutazione su scale, liste di criteri e procedure standardizzate. A ciò si aggiunge un'essenziale preparazione e formazione del/i correttore/i e la presenza di un'equipe costituita da più di un correttore (*inter-rater reliability*).

Le prove soggettive si prestano per lo più a valutazioni qualitative e le prove oggettive a valutazioni quantitative. Oggi però, attraverso l'utilizzo di scale di valutazione dettagliate e condivise, le prove soggettive si avvicinano sempre più a una valutazione quantitativa coerente e attendibile (un esempio concreto è l'utilizzo del QCER).

È possibile, dopo aver assegnato a ogni prova il punteggio grezzo, analizzare e interpretare i dati in vario modo. La valutazione quantitativa può avere luogo mediante procedure statistiche diverse e può fornire differenti tipi di informazione (le informazioni possono riguardare il numero o la percentuale di candidati che hanno risposto correttamente, oppure il numero o la percentuale con cui una particolare risposta è stata scelta dai candidati, ecc.). Vi sono diversi modi per quantificare i dati che dipendono dal tipo di scala di cui ci si avvale.

Le scale di valutazione si possono distinguere, rispettivamente dalla più semplice alla più complessa, in quattro tipologie (nominale, ordinale, a intervalli e di rapporto) corrispondenti a diversi metodi di misurazione. A volte le scale di misurazione sono anche definite “livelli di misurazione”, poiché forniscono quantità di informazioni differenti.

La “scala nominale” prevede la classificazione di elementi o persone mediante l'assegnazione in categorie diverse in base a una particolare caratteristica che le contraddistingue. Ogni elemento classificato può appartenere a una sola categoria ed è omogeneo agli altri elementi appartenenti alla stessa categoria. Non vi sono rapporti di maggioranza, minoranza, superiorità o inferiorità tra i vari elementi. Tale tipologia di scala non consente però, nel caso di test linguistici, di precisare i differenti livelli o gradi di padronanza di una competenza tra i vari soggetti classificati in una stessa categoria. Eppure se si circoscrive l'ampiezza delle qualità o si amplia il numero delle categorie, si può

aumentare il grado di precisazione. I procedimenti consentiti in questa scala sono il calcolo della frequenza e della moda (entrambi sono illustrati nel terzo capitolo). In ambito valutativo un tipo particolare di scala nominale è la scala dicotomica, i cui risultati corrispondono alle risposte vero o falso dei candidati, alle quali si assegna rispettivamente il punteggio pari a 1 e 0.

La “scala ordinale” prevede la classificazione di elementi o persone secondo un ordine di grandezza progressivo. Ogni elemento è classificato secondo un rapporto di maggioranza o minoranza rispetto agli altri elementi. Non è possibile però stabilire la distanza tra una collocazione e l'altra, poiché tra una posizione e l'altra possono esservi intervalli diversi. Il procedimento consentito in questa scala è il calcolo della mediana (illustrato nel prossimo capitolo). In ambito valutativo questo tipo di scala può essere impiegato per le valutazioni scolastiche, in cui si assegnano numeri o giudizi per etichettare una determinata qualità e si costituiscono graduatorie di merito.

La “scala a intervalli” prevede la classificazione di elementi o persone secondo un ordine di grandezza progressivo come per la scala ordinale; però si distingue da quest'ultima perché gli intervalli tra una posizione e un'altra sono sempre uguali e costanti. L'intervallo diviene così anche un esatto indicatore quantitativo, oltre che qualitativo, della distanza tra le due estremità. I procedimenti consentiti in questa scala sono il calcolo della frequenza, della moda, della mediana, della media aritmetica e della deviazione standard (ulteriori chiarimenti nel prossimo capitolo). Questa scala, in ambito valutativo, può essere utilizzata per le prove oggettive, in cui si assegna a priori un peso a ciascun tipo di *item*, determinando così il punteggio massimo conseguibile. Tale punteggio rappresenta l'ampiezza di quella determinata scala per quel determinato test.

La “scala di rapporto” prevede la classificazione di elementi o persone

secondo un ordine di grandezza progressivo, il cui l'intervallo tra le posizioni è sempre uguale e regolare. A differenza della scala a intervallo, questa scala inizia convenzionalmente sempre dal punto zero, che rappresenta l'assenza della caratteristica misurata. Le scale di questo tipo sono utilizzate spesso per misurare caratteristiche fisiche (altezza, peso, ecc.), non trovano invece alcuna applicazione negli studi umanistici.

Queste quattro scale si dispongono secondo un rapporto a senso unico, ossia la scala più complessa può essere convertita in una più semplice, ma non il contrario. È ovvio che la conversione equivale a una perdita di informazioni, ossia di quelle informazioni aggiuntive fornite dalla scala di livello superiore.

Generalmente nell'ambito del *testing* vengono impiegate le scale ordinali e a intervalli. Le scale non sono strumenti di misurazione in senso stretto, in linea di massima si utilizzano per individuare le misure di *tendenza centrale* del fenomeno in esame, ossia misure che permettono di conoscere la tendenza complessiva dei risultati.

2.10 Riflessioni conclusive

Esaminare e valutare sono due processi che implicano la produzione di inferenze sulle competenze linguistiche di un candidato attraverso l'analisi di risultati osservabili.

Alla luce dell'importanza che oggi la valutazione e le certificazioni hanno assunto, sono state illustrate in questo capitolo alcune teorie alla base della valutazione della competenza linguistica e alcuni approcci alla progettazione di *item*. È emerso, attraverso l'analisi del modello di Bachman, che il test è una

metodo per ottenere un certo comportamento da parte del candidato, atteggiamento che viene preso come esempio per inferire su alcune caratteristiche di un certo individuo. Pertanto le caratteristiche del test devono avere delle analogie con la vita reale, in modo da far scaturire un comportamento simile a quello che il candidato avrebbe in un contesto situazionale concreto.

Innanzitutto, per destare la *performance* auspicata, ci si deve soffermare principalmente sulla natura del test, sulla sua costituzione e sullo scopo per cui è progettato. Dunque sono state presentate le fasi di sviluppo di un test, tenendo ben a mente, a ogni passo, lo scopo e l'utilità dello stesso, e facendo attenzione ad alcuni elementi che possono intervenire in una delle fasi, alterando così l'esito della *performance*. Gli effetti di tali elementi, cause potenziali di errori di misurazione, possono essere ridotti, durante la progettazione del test, mediante un'accurata conoscenza degli stessi e delle loro manifestazioni.

Infine si è passati all'analisi dei risultati di un test, il quale deve dar prova di validità, secondo il modello di Weir. Le fasi del processo di misurazione costituiscono la base per l'esame e la dimostrazione dell'affidabilità dei risultati e della validità delle interpretazioni. L'affidabilità dipende dal modo in cui vengono definite e seguite le procedure di misurazione e dal modo in cui i dati ottenuti vengono quantificati. La validità dipende dalla chiarezza con cui sono definiti i costrutti e dall'appropriatezza delle procedure (criteri) scelte per ottenere i risultati attesi. Sebbene la validità possa sembrare la caratteristica più importante, l'affidabilità è apparsa una condizione necessaria per la validità. La validità e l'affidabilità delle prove sono, dunque, entrambi requisiti indispensabili per la considerazione delle stesse come qualità indispensabili per la valutazione della qualità del contenuto di un test e delle competenze linguistiche di un candidato.

Dunque appare evidente come la fase di misurazione e quantificazione

dei risultati sia rilevante per due ragioni: prima di tutto per controllare l'utilità del test prima di considerarlo concluso dal punto di vista dell'implementazione e pronto per l'uso; in seconda istanza per dar prova ai candidati della loro competenza e della validità dei risultati e dei metodi adottati per ottenerli.

Dal momento che la valutazione non si presenta come un processo semplice ed esclusivamente deduttivo, ma comporta analisi empiriche, nel capitolo successivo ci occuperemo di alcune pratiche statistiche per la misurazione e l'analisi dei risultati.

Statistica e analisi dei dati

3.1 Premesse introduttive

Il problema della verifica del profitto e della valutazione dell'individuo interessa da tempo una moltitudine di individui coinvolti professionalmente nella ricerca linguistica e nell'insegnamento delle lingue straniere. Numerosi cambiamenti sono stati progressivamente apportati nel campo della linguistica, della linguistica applicata, dell'insegnamento e della valutazione linguistica. L'attenzione verso il *testing* linguistico si è sviluppata di pari passo con il crescente bisogno di rinnovamento metodologico e contenutistico (in termini soprattutto di bisogni degli apprendenti) e con lo sviluppo di iniziative di ricerca e di sperimentazione (l'interesse per questo ambito è testimoniato dal moltiplicarsi di congressi, corsi e pubblicazioni sull'argomento).

Il *testing* linguistico è un settore della linguistica applicata che, come sostiene Alderson¹, congiunge la valutazione, sotto forma di giudizio, relativa al livello e alla natura dell'apprendimento linguistico di un soggetto con i "dati empirici" relativi alla sua *performance* e, mediante inferenze, alle sue competenze.

La questione relativa alla relazione tra contenuto e dati empirici è

¹ "It is an area of applied linguistics that combines the exercise of professional judgment about language, learning, and the nature of the achievement of language learning with empirical data about student's performances and, by inference, their abilities." (Alderson, in Douglas e Chapelle, 1993:46).

alquanto dibattuta e affrontata in modo diverso dagli esperti del *testing* linguistico di tutto il mondo. La tradizione britannica si è preoccupata più che altro della lingua, dei costrutti che devono stare alla base dei test e che sono oggetto di valutazione. Il versante statunitense, al contrario, si è interessato di pratiche psicometriche e statistiche, riconoscendo nel calcolo la quantificazione di una misura, ossia, in questo caso, la misura della competenza linguistica che un test si propone di misurare (basti pensare al riconoscimento che Valette, nel 1977, ha attribuito alle procedure statistiche e ai cambiamenti nelle tecniche di misurazione e valutazione in corso di svolgimento in quegli anni). Anche in Italia i primi interessi sui dati empirici e sulle procedure statistiche risalgono agli stessi anni con gli studi e le ricerche di Gianfranco Porcelli. Ciò nonostante la maggioranza degli interessati rimane scettica. Spesso insegnanti di lingue straniere, educatori e individui vari legati professionalmente al *testing* linguistico considerano la statistica come una disciplina aliena.

L'obiettivo di un test linguistico consiste nel misurare l'abilità di un soggetto sottoposto a esame, ma senza un'operazione di quantificazione non è possibile precisare in quale misura viene verificata una data conoscenza. La quantificazione, come sostiene Bachman (2004:IX), implica i numeri e i numeri implicano la statistica.

In questo capitolo si descriveranno alcuni concetti statistici basilari, che sembrano necessari per la valutazione linguistica. Tali concetti poggiano comunque sulle teorie della linguistica applicata, le quali forniscono le basi per la definizione delle abilità linguistiche e guidano la progettazione del test e dei rispettivi *task*. I *task* riflettono lo scopo per cui il test è disegnato, ossia la valutazione di una data abilità, e la statistica riguarda il rapporto tra i risultati quantitativi ottenuti dai *task* e il loro significato.

Le analisi statistiche condotte sui risultati dei test forniscono prove empiriche sulla qualità dei test stessi, che altrimenti rimarrebbe legata a convinzioni astratte. L'esame dei dati provvede a fornire indizi pratici al fine di migliorare la qualità e l'utilità delle prove, in modo da assicurare la validità e l'affidabilità dei risultati. Infine un'analisi statistica procura informazioni per una corretta interpretazione dei risultati ottenuti dai candidati sottoposti al test, in modo da inferire in maniera più corretta sulle loro competenze.

3.2 Le analisi statistiche

La valutazione di un test fornisce per lo più dati quantitativi, che spesso si presentano in forma di numeri (si pensi alle prove oggettive). La statistica è un insieme di procedure logiche e matematiche che permettono l'analisi di tali dati (quantitativi).

Spesso dai test si ricavano massicce quantità di dati, provenienti da diversi individui, oppure da misurazioni di differenti e molteplici aspetti di un fenomeno, cosicché è necessario riassumerli per una migliore analisi degli stessi. La statistica fornisce strumenti adatti per riassumere tali dati e interpretarli correttamente.

Le analisi statistiche si possono distinguere in due tipologie, in base allo scopo dell'indagine da condurre: la statistica descrittiva si occupa di interpretare i dati provenienti da un particolare gruppo di individui per inferire sulla loro competenza; la statistica inferenziale si occupa di interpretare i dati di un campione² di individui per poter inferire e generalizzare sulla competenza di un

² In statistica, il termine "campione" indica una parte rappresentativa della popolazione, su cui viene

gruppo più ampio di individui o dell'intera popolazione.

Vi sono numerose procedure statistiche per l'analisi dei dati, pertanto, se si vogliono ottenere risultati significativi, è necessario scegliere le procedure statistiche adatte in base alle caratteristiche dei dati e allo scopo dell'indagine.

In questo capitolo si approfondiranno le procedure statistiche di tipo descrittivo, considerato che la loro comprensione sarà utile per la loro successiva applicazione ai dati, presentati nel prossimo capitolo, che permetterà di inferire sulle *performance* dei soli candidati presi in esame.

3.3 La statistica descrittiva

La statistica descrittiva è una branca della statistica molto importante e costituisce la base per la statistica inferenziale. Fornisce essenzialmente descrizioni sulla distribuzione dei punteggi e permette di stabilire relazioni e differenze tra distribuzioni diverse, di valutare l'affidabilità dei risultati e interpretare tali dati in vario modo.

La maggior parte delle volte un test fornisce dati quantitativi diversi che si distribuiscono su un raggio d'azione che va dal più alto al più basso. Ogni punteggio ottenuto costituisce un valore della variabile di distribuzione.

Per facilitare la lettura di tali punteggi è utile ordinarli dal più alto al più basso disponendoli in una tabella decrescente (tab. n. 3.1 - A).

Inoltre può essere utile raggruppare i punteggi in base alla frequenza con cui si verificano. In tal caso è necessario disporre i punteggi su una prima

effettuata un'indagine da cui inferire dati relativi a tutta la popolazione; il termine "popolazione" indica l'intero oggetto di interesse dell'analisi.

colonna della tabella e il numero della frequenza in una seconda colonna in corrispondenza del relativo punteggio (tab. n. 3.1 - B). In tal modo è possibile individuare con facilità il punteggio più alto o il più basso e il punteggio con il più alto numero di occorrenze. Indubbiamente questo metodo torna più utile quando si ha una grossa mole di punteggi.

Talvolta è persino utile raggruppare i punteggi in intervalli. Per ottenere un intervallo si può dividere la distanza (cosiddetta “*score range*”) tra i punteggi (differenza tra il punteggio più alto e il più basso) per un numero arbitrario di intervalli.

Ad esempio, se si prende in esame il punteggio più alto e il più basso della distribuzione presentata nella tabella n. 3.1, e si vogliono stabilire 6 intervalli, si otterranno sei intervalli di 4 punti ciascuno, infatti $(40 - 17) / 6 = 23 / 6 = 3,8$ (tab. n. 3.1 - C).

Punteggio <i>x</i>	Punteggi o <i>x</i>	Frequenza <i>f</i>	Intervallo punteggi <i>x</i>	Frequenza <i>f</i>
40	40	1	37 - 40	2
37	37	1	33 - 36	5
36	36	2	29 - 32	6
36	35	1	25 - 28	0
35	34	1	21 - 24	1
34	33	1	17 - 20	1
33	32	3		
32	31	2		
32	29	1		
32	24	1		
31	17	1		
31				
29				
24				
17				

A: Punteggi ordinati dal più grande al più piccolo

B: Punteggi raggruppati per numero di occorrenza

C: Punteggi raggruppati in intervalli per numero di occorrenza

Tabella n. 3.1 – Esempi di distribuzione dei punteggi

Dalla colonna C della tabella. n. 3.1 si evince che le occorrenze più

frequenti (pari a 6 casi) cadono nell'intervallo tra 29 e 32.

Anche se la rappresentazione dei dati in tabella ne facilita la lettura, per una migliore interpretazione è più opportuna la rappresentazione grafica su un piano cartesiano, la quale fornisce informazioni sulla forma della distribuzione e agevola l'interpretazione dei risultati. Dal raggruppamento per frequenze e per intervalli non è possibile stabilire la forma che può assumere la distribuzione dei punteggi, la quale è utile, come si vedrà, anche per la scelta di quale procedura statistica applicare tra le numerose possibili.

Spesso si assume che i punteggi si distribuiscono secondo una "distribuzione normale" (soprattutto per i test di tipo normativo), rappresentata da una curva a forma di campana, detta "*curva di Gauss*" dal nome del matematico tedesco che nel 1809 ne descrisse le proprietà.

In una rappresentazione grafica si rappresentano convenzionalmente i punteggi sull'asse orizzontale delle ascisse e le frequenze sull'asse verticale delle ordinate.

Indagini statistiche applicate a rilevazioni numericamente elevate, come sostiene Porcelli (1975:111), hanno dimostrato che le frequenze relative a dati antropometrici³ o psicometrici⁴ si distribuiscono secondo la curva di Gauss.

La curva risulta inizialmente crescente, e raggiunto un apice, comincia a decrescere nuovamente, quasi come se le variabili tendessero a distribuirsi normalmente (fig. n. 3.1).

³ L'"antropometria" studia le proporzioni del corpo umano attraverso la misurazione delle sue diverse parti (cfr. De Mauro, 2000, s.v. "antropometria").

⁴ La "psicometria" si occupa della misurazione dell'intensità, della durata e della frequenza dei processi psichici (cfr. De Mauro, 2000, s.v. "psicometria").

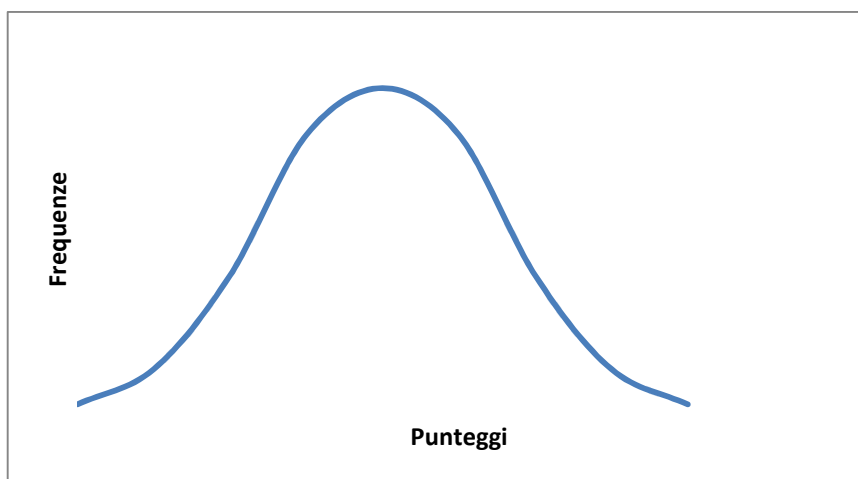


Figura n. 3.1 - Curva di Gauss

Ogni punteggio così viene interpretato in base alla prossimità o alla lontananza dalla normalità distribuitiva. Nella curva di Gauss, la media, la mediana e la moda (questi tre concetti sono spiegati più avanti) coincidono; il valore medio è altresì il più frequente e divide il gruppo in due parti uguali. La curva è simmetrica rispetto al valore medio, quindi quando ci si allontana da esso, sia verso i valori bassi sia verso quelli alti, le frequenze diminuiscono in egual misura.

Se si osserva l'andamento della curva si possono notare alcuni punti in cui la curva cambia direzione, detti "punti di flesso". Questi punti determinano la deviazione standard, ossia la variabilità dalla media complessiva (più avanti viene descritta in dettaglio). Se si riportano sull'ascissa i punti di flesso A e B (fig. n. 3.2) e si rilevano a quali dati rispettivamente A' e B' corrispondono, l'intervallo tra il valore medio M e quello di A' oppure di B' è la deviazione standard.

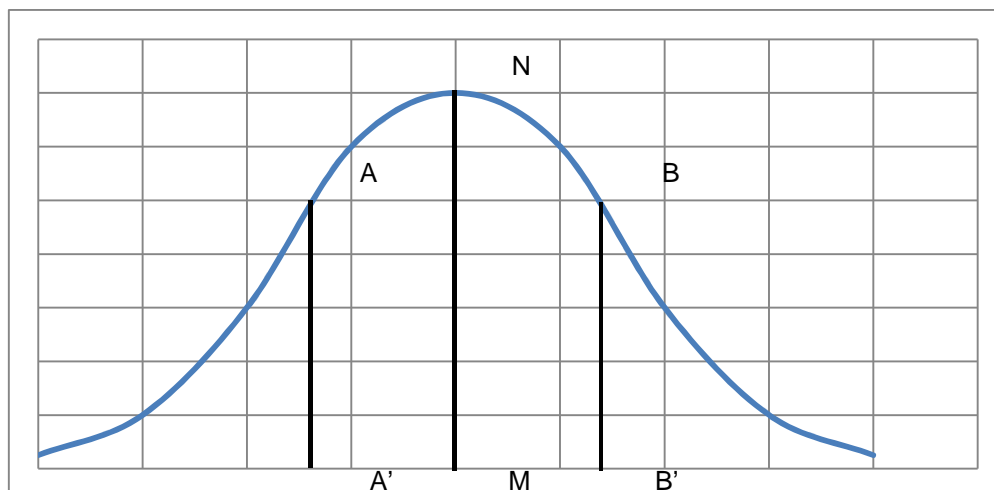


Figura n. 3.2 - La curva di Gauss e la media

È stato calcolato (Porcelli, 1975:113) che nell'area compresa tra AA' e BB' si concentra circa il 68% dei candidati (convenzionalmente si attribuisce a questa porzione centrale il nome di *norma*). Il restante 32% si colloca al di là della norma, un 16% con i dati più bassi e un 16% circa con i dati più alti.

La distribuzione delle frequenze non sempre è normale, ossia simmetrica e a forma di campana. Può, in base alle frequenze, assumere forme diverse, che possono essere simmetriche e avere tuttavia un picco maggiormente allungato verso l'alto (cosiddetta "distribuzione leptocurtica"⁵), o un maggiore appiattimento (cosiddetta "distribuzione platicurtica"); oppure essere asimmetriche e avere un picco all'inizio della distribuzione o alla fine (Bachman, 2004:50-53).

Le rappresentazioni grafiche più diffuse per rappresentare una distribuzione di punteggi sono il "grafico a istogramma" e il "poligono delle frequenze". Nel grafico a istogramma è possibile rappresentare dati di tipo discreto⁶ o di tipo continuo⁷. Il grafico a istogramma mostra, rispettivamente

⁵ La "curtosi" è, in statistica, il grado con cui ci si allontana dalla distribuzione normale, ovvero il maggiore o minore aguzzamento della curva.

⁶ Per dato "discreto" si intende un insieme finito e numerabile di valori possibili.

⁷ Per dato "continuo" si intende un insieme illimitato di tutti i possibili valori di un intervallo.

sull'asse delle ordinate e delle ascisse, la frequenza e i punteggi, sotto forma di barre (separate per i dati discreti, contigue per i dati continui).

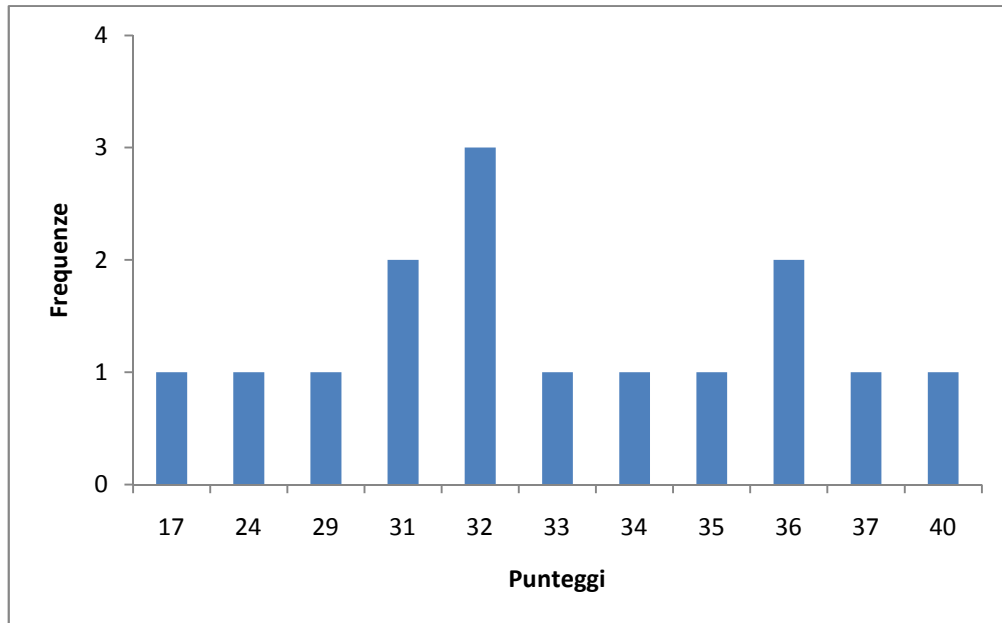


Figura n. 3.3 - Distribuzione dei punteggi in un grafico a istogramma

Per quanto riguarda il poligono delle frequenze, esso può rappresentare solo dati continui. In esso ogni punteggio è rappresentato da un punto, e ogni punto è connesso a un altro da una linea.

Se si vogliono rappresentare i dati della tabella n. 3.1 in un poligono delle frequenze, la rappresentazione si presenta come segue:

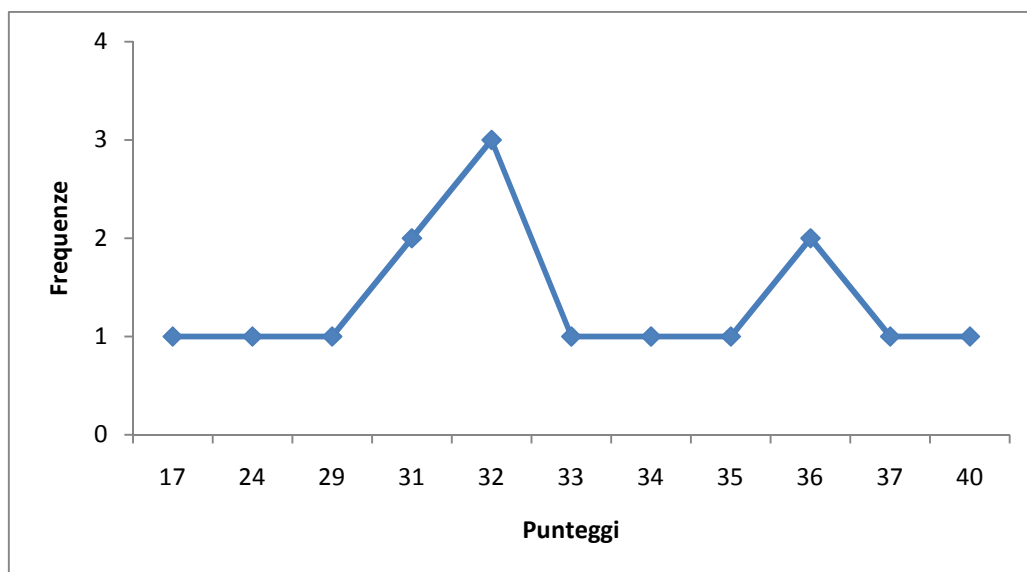


Figura n. 3.4 - Distribuzione dei punteggi in un poligono delle frequenze

La frequenza rappresentata in figura n. 3.4 appare piuttosto spigolosa, ma se si ha a disposizione una grande quantità di punteggi, la rappresentazione della frequenza può assumere una forma più regolare, simile a un poligono a più lati.

Un altro metodo grafico per rappresentare i dati è il “diagramma a scatola a baffi” (*box-and-whisker plot*, comunemente chiamato *box-plot*), in cui è possibile rappresentare il punto centrale della distribuzione e la variabilità dei dati (due concetti statistici che verranno spiegati più avanti: la mediana e lo scarto interquartile).

Il grafico viene rappresentato da un rettangolo, detto “scatola”, delimitato dal primo e dal terzo quartile, diviso in due parti dalla mediana, come nella figura che segue (fig. n. 3.5).

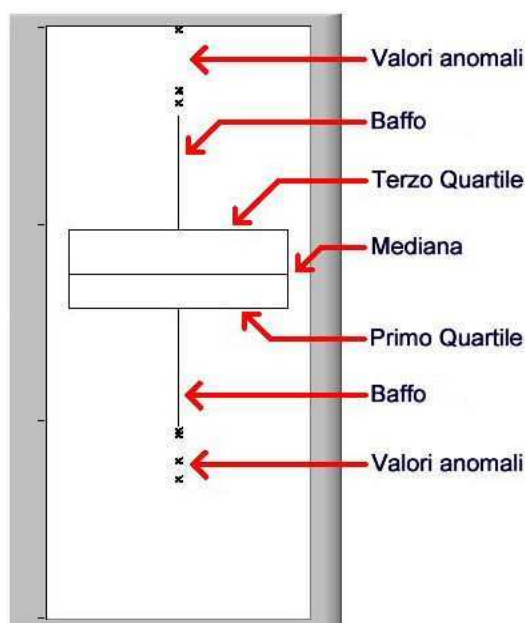


Figura n. 3.5 - Diagramma a scatola

Il quartile, in statistica, corrisponde a ciascuno dei tre valori che ripartiscono una distribuzione in quattro parti uguali (es. in fig. n. 3.6). Il secondo quartile è dunque la mediana, e divide la popolazione in due parti uguali, delle quali il primo quartile (Q_1) e il terzo quartile (Q_3) sono le rispettive mediane.

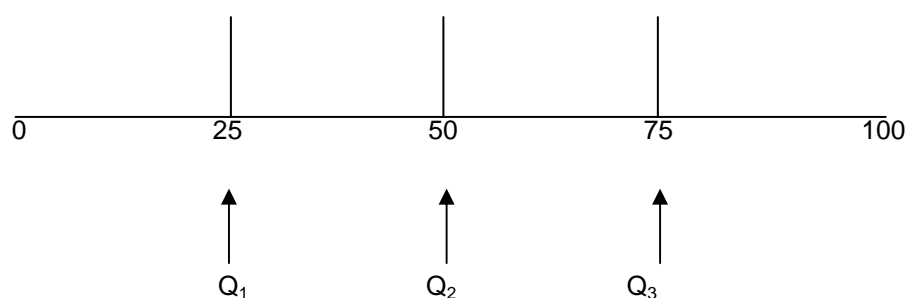


Figura n. 3.6 - Rappresentazione grafica dei quartili

Convenzionalmente, per rappresentare il rettangolo di un *box-plot*, si considerano i tre quartili, di conseguenza l'**intervallo interquartile** è la differenza tra il terzo e il primo quartile ($iqr = Q_3 - Q_1$). L'intervallo interquartile (rappresentato dal rettangolo nella fig. n. 3.5) è la misura della dispersione di una

distribuzione. Il 50% dei dati della distribuzione cade all'interno di questo intervallo. Se l'intervallo interquartile è piccolo, la metà dei dati si trova concentrata intorno alla mediana, se l'intervallo interquartile è grande, la dispersione dei dati centrali intorno alla mediana aumenta.

Le distanze tra ciascun quartile e la mediana forniscono informazioni riguardo alla forma della distribuzione. Nel caso di una distribuzione normale, nel *box-plot* le distanze tra ciascun quartile e la mediana sono uguali, così pure hanno uguale lunghezza le linee che si allungano dai bordi della scatola.

La distribuzione è asimmetrica quando una distanza è diversa dall'altra (fig. n. 3.7).

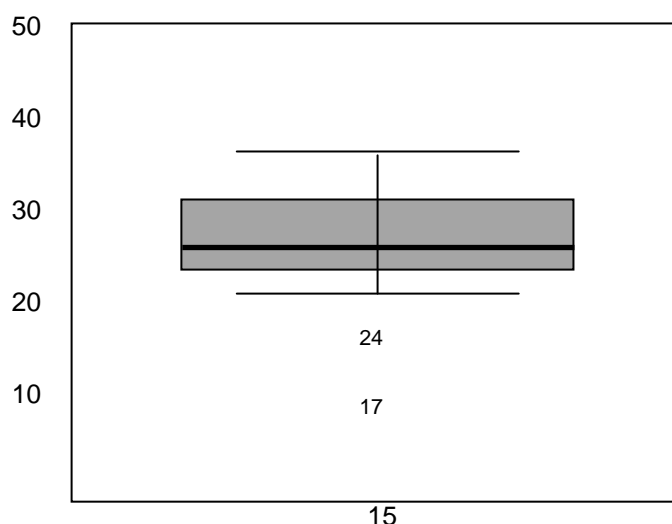


Figura n. 3.7 - Rappresentazione della distribuzione in un *box-plot* (Bachman, 2004:49)

Nella figura n. 3.7 sono rappresentati i dati della tabella n. 3.1: sull'asse delle ordinate vi sono i punteggi e sull'asse delle ascisse le frequenze. L'intervallo interquartile è raffigurato dalla scatola grigia e la linea nera all'interno rappresenta la mediana.

Se i quindici punteggi della tabella n. 3.1 sono divisi dalla mediana in due parti uguali, come nella figura n. 3.8, la mediana dell'intera distribuzione (asse

rosso in fig. n. 3.8) corrisponde al valore 32, il Q_1 corrisponde alla mediana dei valori sull'asse blu (pari a 31) e il Q_3 corrisponde alla mediana dei valori sull'asse verde (pari a 35,5):

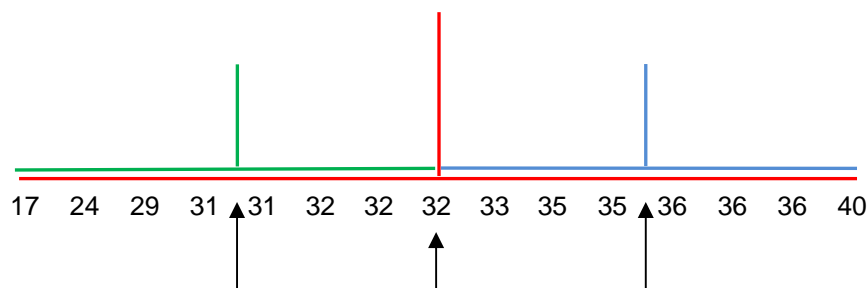


Figura n. 3.8 - Rappresentazione della mediana e dei quartili

Se l'intervallo interquartile è dato dalla differenza tra Q_3 e Q_1 ($35,5 - 31 = 4,5$), il 50% dei dati cade tra 31 e 35,5, in un intervallo di 4,5 punti (31, 31, 32, 32, 32, 33, 35, 35). Al di sotto del primo quartile cade il 25% dei punteggi della distribuzione, al di sopra del terzo quartile cade il restante 25% dei punteggi della distribuzione.

Dal rettangolo fuoriescono due segmenti, detti "baffi". Questi segmenti individuano gli intervalli in cui si dispongono i valori, rispettivamente, minori di Q_1 e maggiori di Q_3 . I punti estremi dei segmenti evidenziano i "valori adiacenti". Il **valore adiacente inferiore** (VAI) è il valore più piccolo della distribuzione e risulta maggiore o uguale al valore ottenuto dalla formula seguente:

$$\text{Formula n. 3.1} \quad \mathbf{VAI = Q_1 - (1,5 \cdot iqr)}$$

Il **valore adiacente superiore** (VAS), invece, è il valore più grande e risulta minore o uguale al risultato della formula seguente:

$$\text{Formula n. 3.2} \quad \mathbf{VAS = Q_3 + (1,5 \cdot iqr)}$$

Se gli estremi della distribuzione sono contenuti tra il VAI e il VAS, essi fungeranno da estremi dei segmenti, altrimenti, come estremi, verranno utilizzati i

valori VAI e VAS.

Si sceglie convenzionalmente di rappresentare tali segmenti in maniera piuttosto ridotta per evitare valori troppo estremi, i quali invece vengono solitamente rappresentati come dei punti esterni rispetto ai valori adiacenti, e sono detti “valori anomali”. Questi valori costituiscono di fatto un’anomalia rispetto alla maggior parte dei dati osservati e pertanto diviene utile identificarli singolarmente.

Nel nostro caso (fig. n. 3.7) il VAI si calcola come: $[31 - (1,5 \cdot 4,5)] = [31 - 6,75] = 24,25$; e il VAS come: $[35,5 + (1,5 \cdot 4,5)] = [35,5 + 6,75] = 42,25$. Il punteggio più alto della distribuzione, pari a 40, risulta dunque inferiore al VAS appena calcolato (42,25), pertanto viene assunto come estremo del segmento superiore. Il punteggio più basso della distribuzione, pari a 29, risulta maggiore del VAI (24,25), pertanto viene assunto come estremo del segmento inferiore.

Il *box-plot* è un metodo che consente di evidenziare la forma della dispersione, la simmetria o asimmetria e la presenza di eventuali valori anomali. Tuttavia, come ogni rappresentazione grafica, si rende utile per una rapida e chiara interpretazione dei risultati ma non fornisce dati precisi, ossia indicatori quantitativi delle caratteristiche della distribuzione dei punteggi. A tale difficoltà sopperiscono alcuni calcoli statistici compiuti sui punteggi.

Le procedure statistiche sono necessarie per la quantificazione dei risultati, e devono essere appropriate al livello di misurazione dei punteggi, cioè alla quantità di informazioni che si possiede in base alla scala di riferimento da cui tali informazioni provengono.

3.3.1 Analisi delle performance

Le procedure statistiche vengono generalmente impiegate per il calcolo della tendenza centrale, per il calcolo della variabilità rispetto alla tendenza centrale e per il calcolo del coefficiente di correlazione.

Tendenza centrale

Le funzioni statistiche che provvedono al raggruppamento dei dati e al calcolo della tendenza centrale sono tre (moda, mediana e media). L'appropriatezza di una piuttosto che di un'altra funzione dipende dal livello di misurazione e dalla prossimità della distribuzione alla normalità distributiva.

La **moda** è il valore che presenta il maggiore numero di frequenze. Nell'ambito del *testing*, la moda è il punteggio conseguito dal maggior numero di candidati.

In una successione di punteggi disposti in ordine di grandezza, come si è già visto prima a grandi linee, la **mediana** corrisponde al punteggio che occupa la posizione centrale nella distribuzione. Se il numero dei punteggi è elevato, la mediana si calcola mediante la formula seguente:

Formula n. 3.3:
$$\mathbf{Md} = \frac{(n+1)}{2}$$

in cui n è il numero dei punteggi. Il risultato ottenuto indicherà la posizione della mediana nella distribuzione dei punteggi. Ciò è possibile se il numero dei punteggi è dispari. Nel caso in cui i punteggi sono di numero pari (dunque non vi è un punteggio centrale), la mediana si ottiene mediante la media aritmetica dei due punteggi centrali (formula n. 3.5) e non corrisponde a un punteggio tra quelli presenti nella distribuzione.

Se il numero dei punteggi è elevato, la mediana può essere calcolata

inoltre mediante la formula seguente:

Formula n. 3.4:
$$\mathbf{Md} = \frac{(n/2) + [(n/2)+1]}{2}$$

in cui n è il numero dei punteggi. Il risultato ottenuto da $(n/2)$ e $[(n/2)+1]$ darà la posizione dei due punteggi centrali, la cui media aritmetica indicherà la mediana.

La **media aritmetica** fornisce un'informazione sull'andamento del test. Si ottiene dividendo la somma (Σ) dei punteggi (x) per il numero (N) dei punteggi stessi, secondo la seguente formula:

Formula n. 3.5:
$$\mathbf{M a} = \frac{\Sigma x}{N}$$

L'esito della formula consente di osservare quanti e quali candidati risultano superiori alla media in una data *performance* e quali e quanti invece risultano inferiori alla media per la stessa *performance*.

Quando si compiono analisi valutative di dati nasce l'esigenza di mettere in relazione le misure più importanti, perché ogni singola misura isolata non fornisce indicazioni complete sui risultati di una *performance*. Quando la mediana è superiore alla media, significa che almeno la metà più uno dei candidati ha conseguito punteggi più elevati rispetto al punteggio medio. Ad esempio, se si hanno i punteggi 4, 5, 7, 8, 9; la mediana è uguale a 7 e la media aritmetica è uguale a 6,6. Quindi tre di cinque candidati hanno ottenuto un punteggio superiore alla media. Quando, al contrario, la mediana è inferiore alla media, significa che la maggior parte dei candidati ha conseguito punteggi inferiori rispetto al punteggio medio. Se, per esempio, si hanno i punteggi 4, 5, 5, 8, 9; la mediana è uguale a 5 e la media aritmetica è uguale a 6,2. Quindi tre di cinque candidati hanno ottenuto un punteggio inferiore alla media.

Quando si ha una grossa quantità di punteggi, di cui molti ricorrono più

volte, è più utile calcolare la media ponderata piuttosto che la media aritmetica. La **media ponderata** è il risultato che si ottiene dalla somma dei singoli valori, che vanno prima moltiplicati con il *peso* (laddove per peso in genere si intende la *frequenza* con cui un valore compare in una sequenza numerica), divisa successivamente per il numero dei punteggi, secondo la formula che segue:

Formula n. 3.6:
$$\mathbf{M a p o n d} = \frac{\sum fx}{N}$$

in cui la somma (Σ) dei prodotti dei valori dei punteggi (x) moltiplicati per la loro frequenza (f) è divisa per il numero (N) dei punteggi.

Finora sono state analizzate misure di tendenza centrale che forniscono informazioni sull'andamento complessivo delle *performance* dei candidati. L'uso appropriato di ciascuna procedura è determinato dalla scala di misurazione. La moda è un indicatore appropriato per una scala nominale, in cui spesso le variabili sono divise per categorie numerate e si rappresenta con la moda la categoria con la percentuale di frequenza più alta. La mediana è un indicatore appropriato per una scala ordinale, in cui si individua l'elemento centrale tra le variabili di una distribuzione. Essendo scarsamente colpita dai punteggi estremi, la mediana, a differenza della media, è un validissimo indicatore della tendenza centrale. Per contro la media è un indice più stabile rispetto alla mediana, e inoltre considera anche la distanza tra i valori, per cui è un valido indicatore della tendenza centrale per ciò che riguarda la scala a intervalli.

Variabilità

Se si vogliono analizzare, al contrario, le variazioni di *performance*, ci si può avvalere di misure di dispersione o variabilità. Quantificare la variabilità significa calcolare la diversità tra due particolari valori della distribuzione o la diversità tra ciascun termine della distribuzione e il valore centrale.

Un indicatore di variabilità è il **campo di variazione** (il *range score* accennato prima), ossia la differenza tra il valore massimo e il valore minimo osservati:

Formula n. 3.7:
$$\mathbf{range} = X_{\max} - X_{\min}$$

Questo indicatore risulta poco descrittivo della distribuzione, in quanto tiene conto solo dei valori estremi, senza tener conto di quelli intermedi.

Quando i punteggi estremi di una distribuzione sono particolarmente lontani dal resto, si è soliti scartarli e calcolare la variabilità solo attraverso i punteggi della tendenza centrale. Si calcola così la differenza tra i punteggi estremi della parte centrale piuttosto che dell'intera distribuzione. In tal caso si calcola la **differenza interquartile** (o intervallo interquartile come accennato prima), ovvero la differenza tra il terzo e il primo quartile della distribuzione:

Formula n. 3.8:
$$\mathbf{iqr} = Q_3 - Q_1$$

A volte si considera la **semidifferenza interquartile**, che è data dalla divisione a metà della differenza interquartile:

Formula n. 3.9:
$$\mathbf{q} = \frac{Q_3 - Q_1}{2}$$

Un indicatore di variabilità che considera tutti i dati, anche i punteggi estremi, esaminandoli singolarmente, è lo **scarto** o **deviazione dalla media**, ossia la differenza tra uno dei punteggi e la media della distribuzione. Lo scarto fornisce un'indicazione di quanto un punteggio differisce dalla media. Tale scarto può avere segno positivo o negativo, a seconda che un dato punteggio si collochi a destra o a sinistra rispetto alla media.

Qualora si volesse calcolare la media degli scarti, ossia la tendenza di variabilità di un gruppo di punteggi, ci si accorgerebbe che la media degli scarti è

sempre uguale a zero. Pertanto è possibile calcolare lo scarto di un punteggio, ma non la misura complessiva della dispersione di tutti i dati. A ciò si può ovviare mediante il calcolo dello **scarto medio**, con cui non si prende in considerazione il segno (positivo o negativo) di ogni singolo scarto, bensì si utilizza il suo valore assoluto. Si calcola la media dei valori assoluti degli scarti e si ottiene un indice di dispersione dei dati:

Formula n. 3.10:
$$SM = \frac{\sum (x - M)}{N}$$

Tale dato viene generalmente utilizzato come base per il calcolo dello scarto quadratico medio. Lo **scarto quadratico medio** o **deviazione standard** indica il grado di omogeneità o eterogeneità dei risultati, quanto cioè i punteggi mediamente si allontanano dalla media complessiva. Si calcola tramite la seguente formula:

Formula n. 3.11:
$$DS = \sqrt{\frac{\sum (x - M)^2}{N}}$$

ossia:

1. si calcola la media aritmetica (M) dei punteggi;
2. si sottrae la media aritmetica da ciascun punteggio (x), per ottenere gli scostamenti ($x - M$);
3. si elevano al quadrato gli scostamenti $(x - M)^2$;
4. si sommano i quadrati degli scostamenti $\sum (x - M)^2$;
5. si divide la somma ricavata per il numero dei punteggi (N);
6. si estraprae infine la radice quadrata ($\sqrt{\quad}$) del valore ricavato, che costituisce la deviazione standard (DS).

Questa procedura si applica quando tutti i punteggi ricorrono con una frequenza pari a 1. Al contrario, se i punteggi ricorrono con più frequenze, prima di sommare i quadrati degli scostamenti (punto 4), è necessario moltiplicare ciascuno di essi per la rispettiva frequenza.

È possibile rappresentare graficamente quanto appena descritto in un

poligono delle frequenze⁸ (fig. n. 3.9), in cui sull'asse delle ascisse si riportano i punteggi e sull'asse delle ordinate si riportano le frequenze.

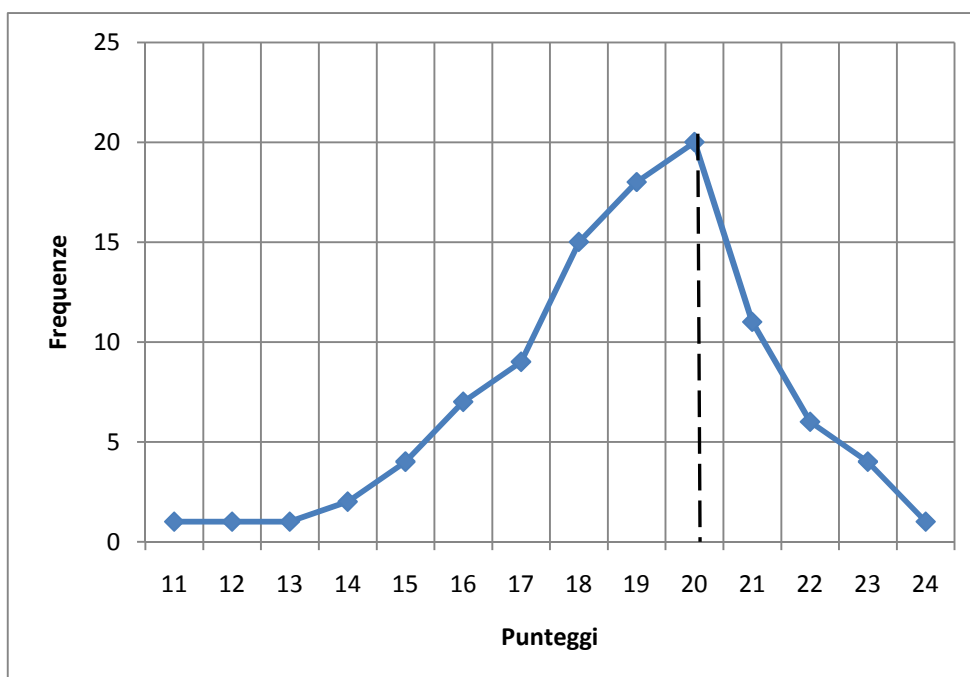


Figura n. 3.9 - Poligono delle frequenze (Porcelli, 1975:111)

La linea tratteggiata rappresenta la media ponderata calcolata su un campione di cento candidati a cui si presume di somministrare un test.

Disposti, ad esempio, in una tabella i dati relativi ai punteggi (x) nella prima colonna, le frequenze (f) con cui ricorre ogni punteggio nella seconda colonna e i prodotti dei punteggi per le rispettive frequenze (fx) nella terza colonna, data la formula n. 3.6 per la media ponderata, si ha che la somma dei singoli valori, moltiplicati prima per la loro frequenza, è uguale a 1880, e che divisa per 100 (il numero dei punteggi), restituisce una media ponderata uguale a 18,8 (tab. n. 3.2).

⁸ Per ulteriori approfondimenti si rimanda a Porcelli, 1975:107-125.

x	f	fx
24	1	24
23	4	92
22	6	132
21	11	231
20	20	400
19	18	342
18	15	270
17	9	153
16	7	112
15	4	60
14	2	28
13	1	13
12	1	12
11	1	11
	N = 100	Σ = 1880

Tabella n. 3.2 - Esempio di rappresentazione delle frequenze

La mediana è data dalla media aritmetica dei due valori centrali, dunque, data la formula n. 3.4 per il calcolo della mediana, si ha $Md = \{((100/2) + [(100/2)+1]) / 2 = Md = \{50 + [50+1]\} / 2 = Md = \{50 + 51\} / 2 =$ (in entrambe le posizioni 50 e 51 è presente un punteggio pari a 19, quindi $Md = \{19+19\} / 2 = 38 / 2 = 19$.

Pertanto sia il grafico che i dati ottenuti dall'applicazione della formula per il calcolo della media e mediana (rispettivamente n. 3.6 e n. 3.4) mostrano come la maggioranza dei candidati abbia conseguito un punteggio superiore alla media; ne è conferma il fatto che il valore della mediana è più alto del valore della media ($Md = 19$; $M a pond = 18,8$).

Per calcolare la deviazione standard dei dati utilizzati nella tabella n. 3.2, data la formula n. 3.11 per il calcolo della deviazione standard, si calcola la media ($M = 18,8$), si calcola la differenza tra ogni punteggio e la media ($x-M$), si elevano al quadrato gli scarti di ogni punteggio dalla media $(x-M)^2$, si moltiplica ciascuno di essi per la rispettiva frequenza $f(x-M)^2$, si sommano tali prodotti ($\Sigma = 580$), si dividono per il numero dei punteggi ($N = 100$), infine si estrae la radice

quadrata e si ottiene una DS = $\sqrt{5,80} = 2,41$ (tab. n. 3.3).

x	f	fx	x-M	(x-M) ²	f (x-M) ²
24	1	24	+5,2	27,04	27,04
23	4	92	+4,2	17,64	70,56
22	6	132	+3,2	10,24	61,44
21	11	231	+2,2	4,84	53,24
20	20	400	+1,2	1,44	28,80
19	18	342	+0,2	0,04	0,72
18	15	270	-0,8	0,64	9,60
17	9	153	-1,8	3,24	29,16
16	7	112	-2,8	7,84	54,88
15	4	60	-3,8	14,44	57,76
14	2	28	-4,8	23,04	46,08
13	1	13	-5,8	33,64	33,64
12	1	12	-6,8	46,24	46,24
11	1	11	-7,8	60,84	60,84
	N = 100	Σ = 1880			Σ = 580,00

Tabella n. 3.3 - Esempio di rappresentazione dei dati per il calcolo della deviazione standard

Questo dato, insieme alla media, risulta utile per stimare il valore di ciascun punteggio. Se si sottrae la deviazione standard dalla media, quindi (18,8 - 2,41) si ottiene un valore pari a 16,39. Secondo le percentuali calcolate sulla base della curva di Gauss, sotto tale punteggio si dovrebbe trovare circa il 16% dei candidati: infatti se si conta il numero dei candidati che ha ottenuto un punteggio inferiore a 16,39, tale numero corrisponde a 16 (su un totale di 100). Se si somma la deviazione standard alla media, quindi (18,8 + 2,41), si ottiene un valore pari a 21,21. Sopra tale punteggio si dovrebbe trovare circa il 16% dei candidati: se si contano i candidati che hanno ottenuto un punteggio superiore a 21,21, essi corrispondono a 11 (su un totale di 100). Si può notare dunque come vi sia una rilevante concordanza tra i dati teorici derivati dalla curva di Gauss e i dati reali.

I punteggi grezzi ottenuti attraverso una misurazione possono risultare di difficile interpretazione. Per rendere più trasparente l'interpretazione dei punteggi grezzi è necessario servirsi di parametri standardizzati e confrontare i punteggi grezzi con una distribuzione di punteggi nota, quale la distribuzione normale. La

procedura statistica che contribuisce a una corretta e condivisa interpretazione dei punteggi è la trasformazione del punteggio grezzo in **punteggio standardizzato** o **punteggio Z**.

Con punteggio Z si identifica la distanza di un punteggio dalla media utilizzando una scala che ha come unità di misura la deviazione standard, come calcolato di seguito:

$$\text{Formula n. 3.12:} \quad Z = \frac{X-M}{DS}$$

in cui X è il punteggio grezzo, M la media, diviso per la deviazione standard DS .

Il punteggio Z consente di uniformare su un'unica scala punteggi provenienti da prove basate su scale diverse, infatti a ogni punteggio grezzo corrisponde un preciso punteggio Z .

È evidente che, qualora il punteggio X sia inferiore alla media M , il punteggio Z avrà valore negativo. Per ovviare a tale inconveniente si può ricorrere alla variante T , trasformando il punteggio Z in **punteggio T** come segue:

$$\text{Formula n. 3.13:} \quad T = 50 + \left(10 \frac{(X-M)}{DS}\right)$$

oppure
$$T = 50 + (10 Z)$$

Per determinare il punteggio T si assegna convenzionalmente alla media un valore uguale a 50 e alla deviazione standard un valore pari a 10. Il risultato della formula n. 3.13 permetterà di posizionare ogni punteggio calcolato con uno scostamento dalla media al di sotto o al di sopra del 50.

Un ulteriore indicatore di variabilità è la **varianza** o **quadrato medio**, ovvero la misura di quanto i dati siano distanti dalla loro media. La varianza è pari alla media aritmetica dei quadrati degli scarti dei valori osservati dalla loro media aritmetica, e si ottiene dalla somma dei quadrati degli scarti dei valori osservati

dalla loro media divisa per il numero di osservazioni, secondo la formula seguente:

Formula n. 3.14:
$$Va = \frac{\sum (x - M)^2}{N}$$

La varianza e la deviazione standard sono direttamente connesse, poiché in pratica la deviazione standard è la radice quadrata della varianza. In realtà questo indice non è molto utile per la statistica descrittiva, poiché è una grandezza quadratica e dunque non direttamente confrontabile con la media o con gli altri valori di una stessa distribuzione. Per questo motivo è preferibile impiegare la deviazione standard, la radice quadrata della varianza permette di trovare una misura espressa nella stessa unità di misura della variabile iniziale (la varianza è invece utilizzata nella statistica inferenziale).

Le procedure per la misurazione della tendenza centrale o della variabilità dei dati non sopperiscono all'esigenza di avere una descrizione della forma di distribuzione. Si è visto come graficamente è possibile stabilire una forma di distribuzione in maniera approssimativa e generica. Tuttavia si può ricorrere a due formule statistiche per ottenere un dato quantitativo della forma di distribuzione: l'indice di curtosi e il coefficiente di asimmetria. Le due formule risultano piuttosto complesse e generalmente vengono calcolate attraverso programmi di statistica, come ad esempio SPSS⁹. Generalmente si considera che si ha un **indice di curtosi** pari a 0 in caso di una distribuzione normale, un indice di numero positivo per un picco maggiormente allungato verso l'alto, e un indice di numero negativo per una distribuzione piatta (tab. n. 3.4). Per il **coefficiente di asimmetria** si considera di solito un valore pari a 0 per una

⁹ SPSS (*Statistical Package for Social Science*) è il software di statistica più diffuso al mondo sia in ambito accademico e di ricerca che in ambito governativo e aziendale. Si tratta di un sistema completo per l'analisi dei dati, consente di elaborare dati provenienti da quasi tutti i tipi di file e di utilizzarli per generare rapporti con tabelle e/o grafici, ed effettuare analisi statistiche avanzate.

distribuzione perfettamente simmetrica (tab. n. 3.4), un valore maggiore di 0 per una distribuzione positiva (su una rappresentazione grafica il picco si forma a sinistra), e un valore minore di 0 per una distribuzione negativa (su una rappresentazione grafica il picco si forma a destra).

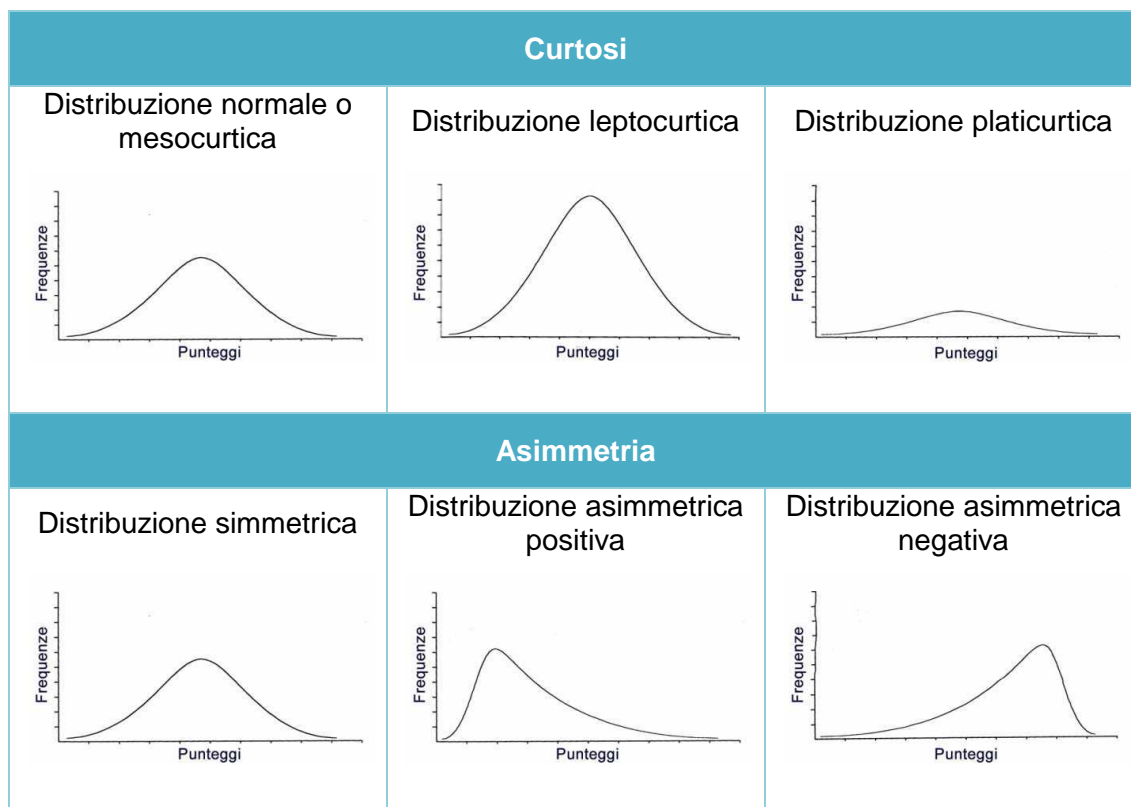


Tabella n. 3.4 - Esempi di distribuzione sulla base dell'indice di curtosi e del coefficiente di asimmetria

Convenzionalmente si assume che sia l'indice di curtosi che il coefficiente di asimmetria compresi tra -2,00 e +2,00 indicano una distribuzione piuttosto normale.

Le procedure di misurazione della variabilità forniscono informazioni sullo scostamento dei valori dalla media. L'uso appropriato di ciascuna procedura è determinato dalla forma di distribuzione dei punteggi e dalla scala di misurazione. Il *range* è un indicatore adeguato di variabilità qualora i dati siano numeri per la descrizione di categorie in una scala nominale (per il cui *range* sarebbe il numero delle categorie). La differenza interquartile o semidifferenza interquartile

è un indicatore appropriato per dati provenienti da una scala ordinale, in quanto non essendo particolarmente colpita dai punteggi estremi, fornisce un valore di dispersione della maggior parte dei valori della distribuzione (che si posizionano verso il centro, intorno alla mediana). La deviazione standard fornisce indicazioni pertinenti alle scale a intervalli, essa si serve di tutti i valori della distribuzione (a differenza del *range* che necessita del punteggio più alto e del più basso ai due estremi della distribuzione; e della differenza interquartile che utilizza il Q_3 e Q_1), è quindi il dato più stabile e attendibile.

Per poter calcolare l'utilità di un test occorre comprendere le caratteristiche di misurazione di alcune procedure statistiche. La prima necessità sta nel determinare la distribuzione di un insieme di punteggi. Si è visto come le rappresentazioni grafiche siano adeguate per delineare la forma della distribuzione, invece le formule statistiche precisano il grado con cui i punteggi si raggruppano intorno a un valore centrale o il grado con cui si allontanano da esso. Queste procedure forniscono informazioni utili per l'interpretazione dei punteggi e per la valutazione dell'affidabilità e della validità di un test.

I procedimenti sinora analizzati contribuiscono a determinare le caratteristiche di una distribuzione, dunque di una sola misura (distribuzione univariata). È tuttavia frequente possedere dati di diverse distribuzioni (distribuzione bivariata) e voler comprendere come queste si relazionino le une con le altre, ad esempio, per uno stesso candidato.

Per rappresentare graficamente le relazioni tra distribuzioni bivariate ci si avvale del "diagramma a dispersione", un grafico bidimensionale in cui due variabili sono rappresentate una sull'asse delle ascisse e l'altra sull'asse delle ordinate. Si hanno dunque due valori, uno per ogni variabile, per ciascun candidato, e si traccia un segno su un grafico cartesiano in corrispondenza del

punto in cui i due valori si incontrano.

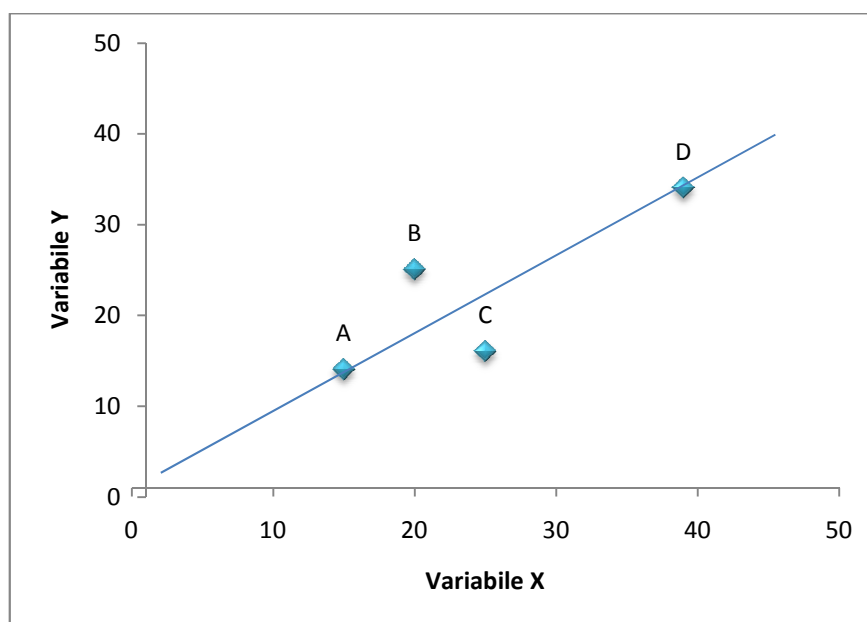


Figura n. 3.10 - Diagramma a dispersione

Nella figura n. 3.10 è possibile notare come vi sia una correlazione diretta, rappresentata da una linea retta, tra due punti, i quali cadono entrambi su di essa. I candidati A e D hanno ottenuto, rispettivamente, il primo punteggi bassi in entrambi i test, il secondo punteggi alti in entrambi i test.

Le relazioni tra due variabili possono variare per potenza e per direzione.

Dal punto di vista della potenza, le relazioni tra le variabili sono forti se i punti tendono a raccogliersi intorno a una linea retta; le relazioni sono deboli se i punti appaiono lontani rispetto alla linea retta. Non vi è invece alcuna relazione se le linee sono tracciabili in ogni angolo del piano cartesiano e persiste la presenza di punti lontani da esse (tab. n. 3.5).

Dal punto di vista della direzione, le relazioni tra le variabili sono dirette o positive se il punteggio alto in una variabile corrisponde a un punteggio altrettanto alto nella seconda variabile; le relazioni sono inverse o negative se il punteggio alto in una variabile corrisponde a un punteggio basso nell'altra

variabile. Nel primo caso si può ipotizzare una linea retta crescente, nel secondo caso una linea retta decrescente (tab. n. 3.5).

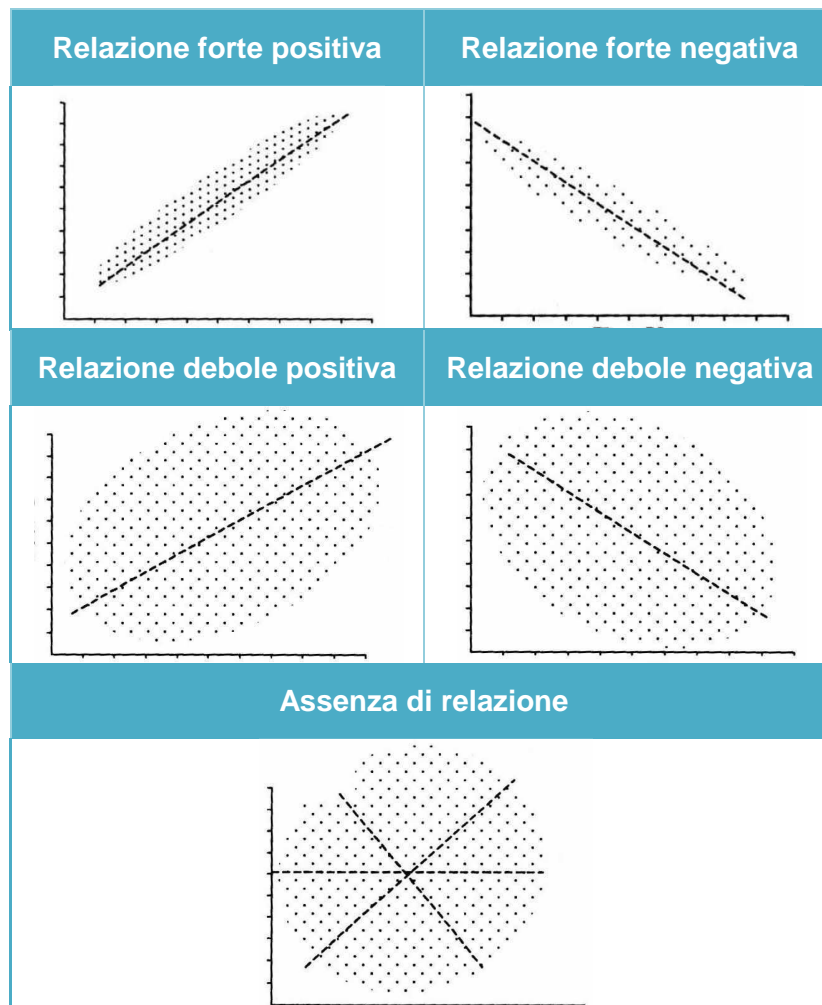


Tabella n. 3.5 - Relazioni tra variabili per potenza e direzione

Tuttavia le rappresentazioni grafiche non procurano dati empirici per l'interpretazione precisa delle relazioni tra due variabili. Per ottenere informazioni sul rapporto tra due variabili occorre calcolare un coefficiente di correlazione.

Coefficiente di correlazione

Per correlazione si intende la tendenza di due variabili a variare insieme, ovvero variano l'una al variare dell'altra, mantenendo lo stesso rapporto di partenza. Quando i punteggi di due distribuzioni diverse si comportano in questo modo si dice che essi tendono a covariare secondo un rapporto lineare. L'indice

che misura la quantità di covariazione tra due variabili è detto **covarianza** e si calcola moltiplicando la distanza di una variabile (X) dalla sua media per la distanza della seconda variabile (Y) dalla sua media, e poi sommando i prodotti ottenuti da ogni coppia di punteggi, secondo la formula seguente:

Formula n. 3.15:
$$\mathbf{cov(xy) = \Sigma(x - M) \cdot (y - M)}$$

Vi sono diversi coefficienti di correlazione, e l'uso appropriato di uno piuttosto che di un altro dipende dal livello di misurazione di ciascuna variabile (tipo di scala), dalla natura della distribuzione (discreta o continua) e dal tipo di distribuzione delle due variabili (lineare o non lineare). Il numero dei coefficienti di correlazione sviluppati dagli statistici è piuttosto alto, qui ci si limita a descriverne due, poiché sono quelli di cui ci si servirà per calcolare alcuni dati nel capitolo successivo.

Un metodo per il calcolo del **coefficiente di correlazione** prende il nome di "*product-moment correlation coefficient*" di **Pearson**. Esso generalmente viene utilizzato per variabili continue misurate con scale a intervalli, che si distribuiscono normalmente e che hanno una relazione lineare. Tale coefficiente (ρ_{xy}) si calcola mediante la seguente formula:

Formula n. 3.16:
$$\rho_{xy} = \frac{N \Sigma_{xy} - (\Sigma_x)(\Sigma_y)}{\sqrt{[N \Sigma x^2 - (\Sigma x)^2][N \Sigma y^2 - (\Sigma y)^2]}}$$

in cui N è il numero dei candidati, Σx è la somma dei punteggi della prima distribuzione, Σy è la somma dei punteggi della seconda distribuzione, Σx^2 è la somma dei quadrati dei punteggi della prima distribuzione, Σy^2 è la somma dei quadrati dei punteggi della seconda distribuzione, Σxy è la somma dei prodotti ottenuti moltiplicando ciascun punteggio x per il punteggio corrispondente y , $(\Sigma x)^2$ è il quadrato della somma delle x , $(\Sigma y)^2$ è il quadrato della somma delle y .

Una formula alternativa per il calcolo del coefficiente di correlazione di

Pearson è la seguente:

Formula n. 3.17:
$$\rho_{xy} = \frac{\sum xy}{(DS_x DS_y)}$$

Essa esprime la linearità tra la covarianza di due variabili ($\sum xy$) derivate da scale a intervalli distribuite normalmente e il prodotto delle rispettive deviazioni standard ($DS_x DS_y$).

Il coefficiente di correlazione di Pearson ha validità unicamente nel caso in cui le variabili sono correlate linearmente e hanno una normalità distributiva.

Il coefficiente di correlazione è espresso da un numero compreso tra +1,00 e -1,00. Una correlazione perfetta è espressa da un indice pari a +1,00 per cui i punteggi delle due distribuzioni risultano correlati positivamente. L'assenza di correlazione è espressa da un indice uguale a 0. Al contrario, un coefficiente pari a -1,00 equivale a una correlazione negativa. Inoltre la correlazione è forte se il coefficiente è maggiore di 0,7; è debole se è compreso tra 0 e 0,3; è moderata se è compreso tra 0,3 e 0,7. Allo stesso modo, per la correlazione negativa, più si avvicina al -1,00 più la correlazione è forte. Il coefficiente di correlazione dunque esprime la variazione di correlazione sia dal punto di vista della potenza che della direzione.

Un altro metodo per il calcolo del coefficiente di correlazione prende il nome di "coefficiente di correlazione per ranghi"¹⁰ (*rank correlation coefficient*) di **Spearman**. Esso si basa sulle differenze di posizione delle variabili nelle due distribuzioni ed è normalmente impiegato per i dati delle scale ordinali che hanno una relazione lineare ma che non devono presentare necessariamente una distribuzione normale.

¹⁰ Per "rango", in statistica, si intende il numero della posizione che una variabile occupa in una distribuzione ordinata.

Dopo aver riordinato i punteggi dal più alto al più basso per ogni distribuzione, si attribuiscono a ciascun punteggio le posizioni in classifica dalla prima all'ultima. Nel caso di più occorrenze dello stesso punteggio, a tale punteggio viene attribuita una posizione, detta "tied rank", pari alla media delle posizioni che si sarebbero attribuite qualora i punteggi fossero stati diversi. Successivamente si confrontano le posizioni tra una graduatoria e l'altra per ogni candidato e si calcolano le differenze. Il calcolo di tale coefficiente, rappresentato dalla lettera greca "rho" (ρ_s), è eseguibile mediante la seguente formula:

Formula n. 3.18:
$$\rho_s = 1 - \frac{6 \cdot \sum D^2}{N \cdot (N^2 - 1)}$$

in cui D indica le differenze di posizione in graduatoria tra una distribuzione e un'altra, N è il numero dei candidati e $\sum D^2$ è la somma dei quadrati delle differenze di posizione.

I valori di correlazione variano, come per il coefficiente di Pearson, tra -1,00 e +1,00. Più i valori sono vicini ai valori +1,00 o -1,00 più la correlazione è forte, più son vicini allo 0 più la correlazione diviene debole. Convenzionalmente si assume che un indice minimo accettabile di correlazione positiva è pari a 0,4.

Spesso si ricorre al coefficiente di correlazione di Pearson quando si hanno grandi quantità di punteggi. Infatti con la formula di Pearson si possono confrontare due serie di dati tra di loro senza ricorrere alla loro collocazione in una graduatoria. In tal modo è possibile comparare, ad esempio, i punteggi di un test (su base 100 o altro) con la tradizionale scala su base 10. Si ricorre invece alla formula di Spearman quando i dati non sono distribuiti normalmente.

Prima di decidere di correlare due variabili, in conclusione, è necessario verificare che ogni distribuzione sia normalmente distribuita. Per far ciò si calcola il valore della curtosi e della asimmetria. Poi, attraverso la rappresentazione

grafica del poligono delle frequenze, si verifica la forma della distribuzione. Se queste procedure dimostrano che la distribuzione è normale, si può calcolare la correlazione mediante sia la formula di Pearson che di Spearman. Se la distribuzione non è normale, è possibile calcolare il coefficiente solo tramite la formula di Spearman.

Il coefficiente di correlazione misura la linearità o la covarianza tra due variabili, tale che a ciascun valore della prima variabile corrisponde con una certa regolarità un valore della seconda variabile. Se si vuol calcolare la proporzione di variabilità di una variabile rispetto all'altra si può calcolare il **coefficiente di determinazione** (r^2), ossia il quadrato del coefficiente di correlazione.

Se invece si vuole predire il valore di una variabile (y) rispetto a un'altra variabile (x) che ha una relazione lineare con la prima, ovvero se si vuol comprendere quanto varia la prima variabile (detta "dipendente") al variare della seconda variabile (detta "indipendente" poiché presa come riferimento per la previsione), si può calcolare il **coefficiente di regressione** mediante la formula che segue:

Formula n. 3.19:
$$\hat{Y} = a + (bX)$$

in cui \hat{Y} è il valore che si prevede per la variabile dipendente, X è il valore della variabile indipendente, a è il segmento o il punto sull'asse delle ordinate in cui inizia la linea retta e b è la pendenza della linea. Con la formula n. 3.19 si vuole predire il valore di Y sulla base del valore di X , cioè si fa "regredire" Y a X .

Ad esempio, se si possiedono i punteggi di due candidati (A, B) relativi a due prove diverse (X, Y), così rappresentati:

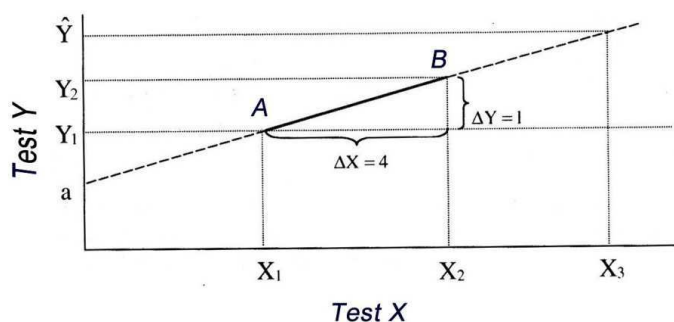


Figura n. 3.11 - Linea di regressione di Y sulla base di X

la linea retta compresa tra A e B rappresenta la relazione tra i punteggi X_1 e Y_1 del candidato A e i punteggi Y_2 e Y_1 del candidato B. La retta, che si estende al di là del punto A e al di là del punto B (linea tratteggiata), estende la regressione al di là dei risultati osservabili.

Per determinare b , ossia la **pendenza della retta**, si utilizza la seguente formula:

Formula n. 3.20:
$$b = \frac{Y_2 - Y_1}{X_2 - X_1}$$

Se si presuppone, ad esempio, che ogni cambiamento Δ è uguale a 1 in Y ($\Delta Y = Y_2 - Y_1 = 1$) e che corrisponde a un cambiamento Δ uguale a 4 in X ($\Delta X = X_2 - X_1 = 4$), si può calcolare la pendenza della retta b come $1/4 = 0,25$.

Per determinare a , ossia il **segmento della retta**, si utilizza la seguente formula:

Formula n. 3.21:
$$a = M_y - (bM_x)$$

in cui M_y è la media della variabile Y, M_x è la media della variabile X, b è la pendenza calcolata in precedenza. Se si presuppone che $M_y = 35$ e $M_x = 48$, si calcola $a = 35 - (0,25 \cdot 48) = 23$.

Tornando alla formula n. 3.19, dopo aver calcolato a e b , e conoscendo X_3 si può calcolare \hat{Y} . Presupponendo che $X_3 = 56$, dato l'esempio precedente, si

calcola $\hat{Y} = 23 + (0,25 \cdot 56) = 37$.

Tale calcolo della regressione è detto calcolo della “regressione lineare” poiché rappresenta un metodo per il calcolo di una variabile dipendente mediante il riferimento a un'altra sola variabile indipendente. Più il numero delle variabili indipendenti a disposizione è alto, più la previsione della variabile dipendente sarà accurata (l'analisi è detta, in tal caso, calcolo della “regressione lineare multivariata”).

I tre coefficienti descritti in questi ultimi paragrafi (di correlazione, determinazione e regressione) sono, dunque, fondamentali per comparare due variabili, per prevedere il valore di una variabile in base al valore di una determinata variabile nota, e per verificare il rapporto previsto fra due variabili.

Riassumendo le formule sinora descritte si potrebbe dire che esse occorrono essenzialmente per quantificare le *performance* dei candidati attraverso il calcolo dei risultati ottenuti nelle prove. Le formule per il calcolo della prossimità a un valore centrale e della dispersione da un valore centrale, e quelle per la correlazione tra due variabili, forniscono informazioni sulle abilità dei candidati. Non va dimenticato però che, quando si progetta un test, lo si fa per un determinato scopo, e generalmente si hanno delle aspettative in merito ai risultati ottenibili (soprattutto se si tratta di prove per verifiche scolastiche, per piccoli gruppi, ecc.). Nel progettare una prova si deve necessariamente tener conto dell'abilità che si vuol misurare e si devono ideare *item* adeguati a essa, che non risultino troppo facili o difficili per i candidati a cui la prova è destinata.

Per tale motivo possono risultare di grande utilità procedure statistiche che consentono di migliorare l'utilità di un test, attraverso l'analisi degli *item*, i quali non devono risultare inefficaci al fine della misurazione. Queste procedure includono l'analisi classica degli *item* che si basa sulla “*Classical Test Theory*”

(CTT), e l'analisi di risposta degli *item* che si basa sulla “*Item Response Theory*” (IRT).

3.3.2 Analisi degli *item*

La teoria classica si basa sull'assunzione che un candidato, che si sottopone a un test, ottiene un “punteggio osservato” che equivale al reale punteggio, ottenuto nella prova, condizionato da un errore casuale, ossia da una componente accidentale non osservabile. I risultati di una prova dovrebbero idealmente riflettere la competenza del candidato, in realtà sono costantemente compromessi da un certo grado di errore. Il tipo di analisi basato sulla CTT si focalizza in genere sull'esame di due caratteristiche dell'*item*: la difficoltà e la discriminazione. Si tratta di descrivere in modo statistico la distribuzione dei punteggi degli *item*. Con le domande che presuppongono risposte di tipo giusto/sbagliato (R/W), attraverso l'analisi degli *item* (IA), si possono identificare i problemi che sorgono con la risposta corretta (*key*) oppure con le risposte errate (distrattori). Con le domande a credito parziale (P-C), ossia domande con risposte esatte, risposte errate e risposte ritenute parzialmente corrette (che permettono al candidato di ottenere un punteggio parziale che va da zero al valore stabilito come massimo punteggio), si possono identificare i problemi che sorgono con i punteggi diversi ottenuti nelle diverse risposte.

La difficoltà di un *item* si ottiene dalla proporzione dei candidati che rispondono in modo corretto a un dato *item*, (qualora si tratti di domande che presuppongono risposte di tipo giusto/sbagliato); oppure si ottiene dal punteggio medio ottenuto nel dato *item* (qualora si tratti di domande a credito parziale). È possibile calcolare la difficoltà di un *item* sia per le prove di tipo normativo che

per quelle di tipo criteriale.

La discriminazione di un *item* fornisce la capacità di un *item* di distinguere i candidati più bravi da quelli meno competenti. È possibile calcolare la discriminazione di un *item* sia per le prove di tipo normativo che per quelle di tipo criteriale.

Per le prove di tipo normativo, per constatare il grado di difficoltà di un *item*, si calcola l'**indice di difficoltà**¹¹ mediante il rapporto tra il numero di risposte esatte (E) e il numero dei candidati (N) che hanno affrontato l'*item*:

Formula n. 3.22:
$$IF = \frac{E}{N}$$

In ultima analisi, con le domande di tipo R/W, che costituiscono una distribuzione dicotomica (1/0), si tratta di calcolare la media così come descritta nella formula n. 3.5, ovvero $IF = \frac{\sum X}{N}$, in cui x è il punteggio di un dato individuo in un dato *item* e N è il numero dei candidati. Se un *item* ottiene un punteggio medio inferiore rispetto a un punteggio medio di un altro *item*, il primo risulta più difficile del secondo.

L'indice è sempre un numero razionale compreso tra 0 (difficoltà uguale a zero) e 1 (difficoltà massima). L'indice di difficoltà è accettabile se risulta compreso tra 0,7 e 0,3, poiché se l'indice è al di sotto del valore 0,3 l'*item* è troppo difficile, se è al di sopra del valore 0,7 l'*item* è troppo facile.

Per calcolare la **percentuale di difficoltà** si può utilizzare la seguente formula:

Formula n. 3.23:
$$IF = \frac{E \cdot 100}{N}$$

¹¹ L'indice di difficoltà è anche detto "indice di facilità". Pertanto per indicare questo indice qui si utilizzerà l'acronimo "IF" in modo da distinguerlo dall'indice di discriminazione "ID".

Un *item* è facile se il 75% o più dei candidati risponde in maniera esatta, è di media difficoltà se l'IF è compreso tra 25% e 75%, è difficile se meno del 25% dei candidati risponde in modo esatto. Se l'IF è uguale al 100% oppure allo 0%, significa che l'*item* presenta una capacità di discriminazione nulla, ossia la sua capacità selettiva o di differenziazione tra le *performance* dei vari candidati è pari a zero. Se in un *item*, sia i candidati migliori che i peggiori danno la stessa percentuale di risposte esatte, il grado di differenziazione è ugualmente nullo.

Se si tratta di calcolare l'indice per un grosso numero di candidati risulta più conveniente disporre le prove in ordine di punteggio, dal più alto al più basso, poi dividerle in tre gruppi, di cui nel primo e nel terzo si avranno per ciascuno il 27% dei candidati che hanno conseguito rispettivamente i punteggi più alti e i punteggi più bassi (ciascun gruppo deve costituirsi necessariamente dello stesso numero di individui). Generalmente si contano poi le diverse risposte per ogni *item* utilizzando una tabella come la seguente:

Item Tally Sheet							
Tipo di prova:		Gruppo:		Totale candidati:		Data	
N. item	Risposte	Gruppo h n. candidati: 20	P _h	Gruppo l n. candidati: 20	P _l	IF	ID
1	a	2	0,1	7	0,35	0,23	- 0,25
	b	18	0,9	8	0,4	0,65	0,5
	c		0,0	4	0,2	0,1	- 0,2
	d		0,0	1	0,05	0,03	- 0,05
	O		0,0		0,0	0,0	0,0
2	a	2	0,1	4	0,2	0,15	- 0,1
	b	4	0,2	10	0,5	0,35	- 0,3
	c	13	0,65	2	0,1	0,38	0,55
	d	1	0,05	4	0,2	0,13	- 0,15
	O		0,0		0,0	0,0	0,0
3	a	1	0,05	2	0,1	0,08	- 0,05
	b	4	0,2	11	0,55	0,38	- 0,35
	c	6	0,3	4	0,2	0,25	0,1
	d	9	0,45	3	0,15	0,3	0,3
	O		0,0		0,0	0,0	0,0

Tabella n. 3.6 - Esempio di conteggio degli *item* (Bachman, 2004:124)¹²

¹² Laddove i valori variano rispetto all'originale, la ragione è legata a operazioni di ricalcolo e

Si inserisce, ad ogni *item*, un segno (✓) per ogni opzione scelta (a, b, c, d, O=omissione) da ciascun candidato, prima per il gruppo con i punteggi più alti (*h*), poi per il gruppo con i punteggi più bassi (*l*). Successivamente si conta il numero dei segni per ogni gruppo per verificare che corrisponda al totale dei candidati appartenenti allo stesso e poi si calcola la proporzione per ogni opzione scelta da un dato numero di candidati diviso il totale dei candidati del gruppo a cui appartengono, e si inserisce il valore ottenuto dalla proporzione nelle colonne p_h e p_l (ad es. l'opzione *a* del primo *item* è stata scelta da 2 candidati del gruppo *h* su 20, dunque $2 / 20 = 0,1$; e da 7 candidati appartenenti al gruppo *l*, dunque $7 / 20 = 0,35$).

A questo punto si può calcolare l'indice di difficoltà secondo la formula che segue:

Formula n. 3.24:
$$IF = \frac{p_h + p_l}{2}$$

in cui p_h è per l'appunto la proporzione dei candidati che hanno risposto in modo corretto tra quelli che appartengono al gruppo che ha ottenuto i punteggi più alti (*h*); e p_l è la proporzione dei candidati che hanno risposto in modo corretto tra quelli che appartengono al gruppo che ha ottenuto i punteggi più bassi (*l*). La risposta corretta per ciascun *item* è contrassegnata, nella tabella n. 3.6, dalla lettera sottolineata nella colonna delle risposte. Ad esempio il calcolo dell'IF del primo *item* è uguale a $IF = (0,9 + 0,4) / 2 = 1,3 / 2 = 0,65$.

La formula n. 3.24 equivale alla seguente formula:

Formula n. 3.25:
$$IF = \frac{R_h + R_l}{2n}$$

in cui R_h è il numero dei candidati appartenenti al gruppo *h* che ha risposto

correzione degli errori riscontrati.

correttamente; R_l è il numero dei candidati appartenenti al gruppo l che ha risposto correttamente; n è il numero dei candidati del gruppo h o l (in entrambi il numero dei candidati è uguale). Ad esempio il calcolo dell'IF del primo *item* con la formula n. 3.25 è uguale a $IF = (18 + 8) / (2 \cdot 20) = 26 / 40 = 0,65$. Le formule n. 3.24 e n. 3.25 sono analoghe.

È possibile calcolare anche la proporzione dei candidati che hanno scelto un distrattore, ossia l'**indice di distrattività** di una risposta errata, mediante la formula:

Formula n. 3.26:
$$p_x = \frac{n_x}{N}$$

in cui n_x è il numero dei candidati che ha scelto un distrattore X ; N è il numero dei candidati.

Oppure, per grandi gruppi, è possibile considerare soltanto il primo (h) e terzo (l) gruppo di candidati applicando la seguente formula:

Formula n. 3.27:
$$p_x = \frac{nr_h + nr_l}{2n}$$

in cui nr_h è il numero dei candidati del gruppo h che ha scelto il distrattore X , nr_l è il numero dei candidati del gruppo l che ha scelto il distrattore X , n è il numero dei candidati del gruppo h o l (in entrambi il numero dei candidati è uguale).

Oppure ancora la seguente formula:

Formula n. 3.28:
$$p_x = \frac{p_h + p_l}{2}$$

in cui p_h è la proporzione dei candidati del gruppo h che hanno scelto il distrattore X , p_l è la proporzione dei candidati del gruppo l che hanno scelto il distrattore X . Ad esempio, per il distrattore b del secondo *item*, p_x è $= (4 + 10) / (2 \cdot 20) = 14 / 40 = 0,35$ applicando la formula n. 3.27; ugualmente, applicando la formula n. 3.28, p_x è $= (0,2 + 0,5) / 2 = 0,7 / 2 = 0,35$. Le formule n. 3.27 e n. 3.28 sono

equivalenti.

Per quanto riguarda l'indice di difficoltà di un distrattore, si assume che il valore deve essere almeno uguale o superiore a 0,1. Se un distrattore non viene scelto da nessun candidato ($p_x = 0$), significa che esso non svolge la sua funzione e va modificato. Convenzionalmente si selezionano *item* con distrattori che hanno un p_x uguale o maggiore di 0,3.

Per stabilire l'indice di difficoltà di un *item* con domande di tipo P-C occorre calcolarne la media. Per gruppi piccoli è possibile prendere in considerazione tutti i test, dunque calcolare l'indice secondo la seguente formula:

Formula n. 3.29:
$$IF = \frac{\sum x}{N}$$

in cui x è il punteggio di un dato individuo in un dato *item* e N è il numero dei candidati.

Per i grossi gruppi è opportuno adottare il metodo dei due gruppi, rispettivamente con i punteggi più alti e con quelli più bassi, dunque l'IF viene calcolato come segue:

Formula n. 3.30:
$$IF = \frac{M_h + M_l}{2}$$

in cui M_h è la media dell'*item* calcolata sul gruppo h , M_l è la media dell'*item* calcolata sul gruppo l . Poiché però l'*item*, in questi casi, può ottenere punteggi scalari (ricordiamo che si tratta di domande di tipo P-C), che raggiungono come punteggio massimo valori arbitrari diversi di volta in volta, è necessario dividere ogni media per il possibile intervallo compreso tra il punteggio massimo e il punteggio minimo:

Formula n. 3.31:
$$IF = \frac{M}{X_{max} - X_{min}}$$

in cui M è la media dell'*item*, X_{max} è il punteggio massimo possibile e X_{min} è il

punteggio minimo possibile per il dato *item*.

Anche per i test di tipo P-C è possibile calcolare la proporzione dei candidati che scelgono un dato distrattore mediante la formula n. 3.28.

Per constatare il grado di differenziazione di un *item* si calcola l'**indice di discriminazione**. Per le domande di tipo R/W si applica la seguente formula:

Formula n. 3.32:
$$ID = \frac{R_h - R_l}{N}$$

in cui R_h è il numero dei candidati appartenenti al gruppo h che ha risposto correttamente; R_l è il numero dei candidati appartenenti al gruppo l che ha risposto correttamente; n è il numero dei candidati che compongono il gruppo h o l (in entrambi il numero dei candidati è uguale). Ad esempio il calcolo dell'ID del primo *item* è $= (18 - 8) / 20 = 10 / 20 = 0,5$.

La formula per il calcolo della proporzione è la seguente:

Formula n. 3.33:
$$ID = (p_h - p_l)$$

in cui p_h è per l'appunto la proporzione dei candidati che hanno risposto in modo corretto tra quelli che appartengono al gruppo h e p_l è la proporzione dei candidati che hanno risposto in modo corretto tra quelli che appartengono al gruppo l . Ad esempio il calcolo dell'ID del primo *item* con la formula n. 3.33 è $= (0,9 - 0,4) = 0,5$.

È possibile calcolare anche l'**indice di discriminazione dei distrattori**, ossia delle risposte errate, mediante la formula che segue:

Formula n. 3.34:
$$p_x = \frac{nr_h - nr_l}{n}$$

in cui nr_h è il numero dei candidati del gruppo h che ha scelto il distrattore X , nr_l è il numero dei candidati del gruppo l che ha scelto il distrattore X , n è il numero dei candidati del gruppo h o l (in entrambi il numero dei candidati è uguale). Ad

esempio, per il distrattore b del secondo *item*, p_x è $= (4 - 10) / 20 = -6 / 20 = -0,3$.

Oppure ancora, è possibile calcolare la proporzione dei candidati che scelgono il distrattore, attraverso la seguente formula:

Formula n. 3.35:
$$p_x = (p_h - p_l)$$

in cui p_h è la proporzione dei candidati del gruppo h che hanno scelto il distrattore X , p_l è la proporzione dei candidati del gruppo l che hanno scelto il distrattore X . Ad esempio, per il distrattore b del secondo *item*, $p_x = (0,2 - 0,5) = -0,3$. Le formule n. 3.34 e n. 3.35 sostanzialmente equivalgono.

È evidente, dal calcolo dell'esempio appena mostrato, che l'ID può avere anche un valore negativo. Ciò vuol dire che l'ID varia da +1,00 (discriminazione massima), in cui tutti i candidati migliori rispondono in modo esatto e tutti i peggiori rispondono in maniera errata, a -1,00 (discriminazione rovesciata massima), in cui tutti i candidati migliori rispondono in maniera errata e tutti i peggiori rispondono in modo esatto. L'indice di discriminazione deve essere compreso tra 0,3 e 0,7, poiché un ID al di sotto di 0,3 indica un *item* non discriminante, al di sopra di 0,7 uno troppo selettivo.

Per le domande di tipo P-C l'indice di discriminazione si calcola attraverso la differenza tra le medie ottenute dai gruppi h e l , diviso l'intervallo compreso tra il punteggio massimo e quello minimo:

Formula n. 3.36:
$$ID = \frac{M_h - M_l}{X_{max} - X_{min}}$$

Per questo tipo di domande l'indice di discriminazione per i distrattori è calcolabile mediante le formule n. 3.34 e n. 3.35.

L'ID, per essere funzionale, ovvero per avere la capacità di differenziare i candidati più meritevoli da quelli meno bravi, deve essere un valore positivo

piuttosto alto, se calcolato per la risposta corretta; se calcolato per il distrattore, deve essere un valore negativo poiché vuol dire che è stato scelto da più candidati appartenenti al gruppo l che al gruppo h (gruppo in cui si presume il distrattore non debba essere scelto da una grande percentuale di candidati).

Per le prove di tipo criteriale l'indice di difficoltà si calcola pressappoco come per le prove di tipo normativo. La differenza consiste nello stabilire *apriori* un livello di competenza da raggiungere, assunto come criterio di valutazione. In genere si ordinano le prove in base al punteggio, dal più alto al più basso, poi si suddividono in due gruppi, da una parte i candidati che hanno raggiunto il livello di padronanza (p) e dall'altra i candidati che non lo hanno raggiunto (np). Successivamente si segnano le risposte in una tabella come quella illustrata nella tabella n. 3.6 e si calcolano gli indici p_p e p_{np} . L'IF, per le domande di tipo R/W, si calcola mediante la formula n. 3.22, mentre per calcolare l'IF di un distrattore si esegue la seguente formula:

$$\text{Formula n. 3.37:} \quad \mathbf{p_x} = \frac{nr_p + nr_{np}}{n_p + n_{np}}$$

in cui nr_p è il numero dei candidati del gruppo p che ha scelto il distrattore X , nr_{np} è il numero dei candidati del gruppo np che ha scelto il distrattore X , n_p è il numero dei candidati del gruppo p e n_{np} è il numero dei candidati del gruppo np .

Per le domande di tipo P-C, per il calcolo dell'IF si impiegano le formule n. 3.29, n. 3.30 e n. 3.31.

L'indice di discriminazione si può calcolare, per le domande di tipo R/W, mediante la seguente formula:

$$\text{Formula n. 3.38:} \quad \mathbf{ID} = \frac{nr_p}{n_p} - \frac{nr_{np}}{n_{np}}$$

E per le domande di tipo P-C con la formula che segue:

Formula n. 3.39:
$$ID = \frac{M_p - M_{np}}{X_{max} - X_{min}}$$

in cui M_p è la media ottenuta nell'*item* dal gruppo p , M_{np} è la media ottenuta dal gruppo np , X_{max} è il punteggio massimo possibile e X_{min} è il punteggio minimo possibile per l'*item*.

Una volta che gli indici di difficoltà e di discriminazione per le prove normative e criteriali sono stati calcolati, è possibile stabilire se un *item* è adeguato per il test, oppure se va rivisto o eliminato. Naturalmente, prima di servirsi di procedure statistiche, è necessario esaminare il contenuto, le specificazioni della prova e lo scopo.

Le procedure statistiche sono un'integrazione all'analisi selettiva degli *item*, affinché sia un'analisi di tipo tanto qualitativo quanto quantitativo. L'analisi qualitativa del contenuto di un *item*, oltre a verificare che l'*item* rispetti lo scopo e le specificazioni del test, deve servire per verificare che non ripeta il contenuto di un altro *item* già selezionato per lo stesso test.

L'analisi degli *item* può risultare limitata per certi aspetti o situazioni. A tal proposito può essere utile riflettere sul fatto che l'analisi degli *item* si basa sulle risposte date da un determinato gruppo, per capire che, molto probabilmente, somministrando lo stesso tipo di prova a un altro gruppo, gli indici di difficoltà e di discriminazione potrebbero presentare valori differenti. Una soluzione potrebbe essere la costituzione di una grande banca dati di *item*, in cui ciascun *item* è classificato secondo le specificazioni del contenuto e secondo gli indici di difficoltà e di discriminazione, basati sulla somministrazione a gruppi diversi.

Un secondo limite della IA è legato all'incapacità di prendere in considerazione il livello di abilità del singolo candidato. Il valore p_h o p_t , che determina la proporzione dei candidati che scelgono una data risposta tra le

opzioni e il totale dei candidati che compongono il gruppo h o l , fornisce informazioni sulla *performance* generale dell'intero gruppo, non fornisce informazioni sulla *performance* di un particolare candidato (il risultato totale di un test fornisce informazioni sulla *performance* generale di un singolo candidato, non sulla *performance* in un determinato *item*).

L'analisi di risposta degli *item* può sopperire a queste limitazioni. Come accennato prima, questo tipo di analisi si basa sull'*Item Response Theory* (IRT). L'IRT supera il concetto di dipendenza tra candidato e *item* della CCT. Secondo l'IRT la *performance* di un candidato è determinata da un tratto latente dell'abilità del candidato e dalle caratteristiche dell'*item*. L'abilità latente del candidato è però indipendente dal tipo di test, ossia il punteggio ottenuto da un individuo non dipende dalla specifica prova somministratagli, e le caratteristiche di un *item* non dipendono dal particolare gruppo di candidati a cui l'*item* è destinato.

Sulla IRT si basano diversi modelli che variano per il numero di elementi presi in considerazione, quali la dimensione, i tratti nascosti e le caratteristiche dell'*item*. Il modello più noto fa riferimento al presupposto della "unidimensionalità", ossia al principio che alla base di una *performance* per un dato *item* risiede un solo tratto o abilità. Il livello di tale abilità, rappresentato dalla lettera greca "*theta*" (θ), viene stabilito sulla base della *performance* del candidato in un particolare set di domande. Inoltre la risposta che il candidato fornisce in un *item* è indipendente dalle risposte fornite negli altri *item* dello stesso test.

Per l'analisi di risposta degli *item* generalmente si prendono in considerazione modelli IRT che si possono applicare in presenza di *item* dicotomici (in cui vi sono solo due possibili risposte R/W o punteggi 0/1).

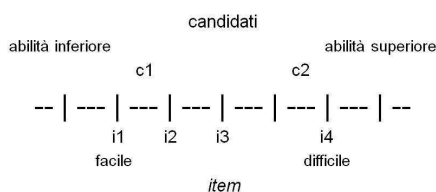
I modelli che si basano sull'IRT variano anche per il numero di parametri

utilizzati per la valutazione statistica della *performance* dei candidati. I modelli più noti sono il modello a un solo parametro (detto anche “modello di Rasch”¹³ dal nome del matematico danese G. W. Rasch), in cui si prende in considerazione solo il livello di difficoltà dell’*item*; il modello a due parametri in cui si prende in considerazione il livello di difficoltà e il livello di discriminazione dell’*item*; il modello a tre parametri in cui si prende in considerazione, in aggiunta ai due parametri appena menzionati, il parametro cosiddetto di “*guessing*”, ovvero l’influenza del caso nel determinare il superamento dell’*item* (la probabilità che i candidati meno competenti possano rispondere correttamente poiché tentano di indovinare). Nel modello a un parametro si assume che il livello di discriminazione è costante per tutti gli *item* mentre il parametro di *guessing* non è assolutamente considerato; in ultima analisi è il modello più semplice e probabilmente quello più utilizzato.

La IRT si basa sull’ipotesi che è possibile prevedere la risposta corretta di un *item* in modo probabilistico sulla base della relazione tra le caratteristiche dell’*item* e le abilità dei candidati. La funzione matematica con cui viene

¹³ Si tratta di un modello matematico che mette in relazione la *performance* di un candidato in un dato test con il suo livello di abilità. Esso si basa sulla teoria secondo la quale la risposta data da un candidato dipende dal livello di difficoltà dell’*item* e dal livello di abilità del candidato stesso. È un modello probabilistico poiché prende in considerazione la probabilità che gli stessi modelli di risposta possano riprodursi in soggetti con abilità e difficoltà simili a quelle di soggetti già noti.

Ad esempio:



il candidato c2 è più abile del candidato c1, pertanto il suo livello gli permette di svolgere gli *item* i1, i2, i3 ma non i4. È quindi probabile che un candidato con un livello uguale al candidato c2 svolga solo gli *item* i1, i2 e i3.

Inoltre il candidato c2 ha più probabilità di superare l’*item* i1 piuttosto che l’*item* i3 poiché è molto vicino al suo livello di abilità.

Per ulteriori chiarimenti sul modello di Rasch si rimanda a Henning, 1987:117-125.

formalizzata tale relazione è detta “*Item Characteristic Curve*” (ICC) per i modelli unidimensionali e “*Item Characteristic Function*” (ICF) per i modelli multidimensionali.

Nel modello di Rasch la **probabilità di rispondere correttamente a un *item*** dipende dal livello di abilità del candidato e dalla difficoltà dell'*item*. La ICC di questo modello si rende evidente con la seguente funzione matematica:

$$\text{Formula n. 3.40:} \quad P_j(\theta_i) = \left(\frac{e^{(\theta_i - b_j)}}{1 + e^{(\theta_i - b_j)}} \right)$$

in cui $P_j(\theta_i)$ indica la probabilità che un candidato (i) risponda correttamente all'*item* (j); e è una costante matematica (detta “costante di Nepero”) rappresentata dal numero irrazionale 2,7182818284 (si tratta dell'unico numero positivo per il quale $e^x \geq 1 + x$); θ_i è l'abilità del candidato e b_j indica la difficoltà dell'*item*. La funzione esponenziale¹⁴, rappresentata nella formula n. 3.40, suggerisce sostanzialmente che la probabilità ($P_j(\theta_i)$) che un candidato (i) risponda correttamente all'*item* (j) dipende dalla differenza tra l'abilità del candidato (θ_i) e la difficoltà dell'*item* (b_j).

Questa differenza ($\theta_i - b_j$) è potenzialmente infinita, quindi, a fronte di una probabilità ($P_j(\theta_i)$) variabile tra 0 e 1, si può applicare la **funzione logit**¹⁵, come

¹⁴ La funzione esponenziale è una funzione del tipo: $y = a^x$, in cui $a > 0$, $x \in R$ (\in = appartiene, R = insieme dei numeri reali). Un'equazione si dice esponenziale quando l'incognita compare soltanto nell'esponente di una potenza, ad es. $a^x = b$.

Il logaritmo è la funzione inversa dell'esponenziale: $y = \log_a x$, in cui $a > 0$, $x \in R^+$. L'inverso dell'esempio precedente è: $x = \log_a b$ (a è la base di entrambe le funzioni). Dunque il logaritmo di un numero secondo una certa base è quel numero per cui si deve elevare quella base per ottenere il numero dato. Ad es. 4 è il logaritmo in base 3 di 81: $\log_3 81 = 4$; infatti l'operazione inversa, ossia l'elevamento a potenza, è uguale a $3^4 = 81$.

Per ulteriori chiarimenti sui concetti di funzione esponenziale e logaritmo fare riferimento a un qualunque manuale di matematica o statistica.

¹⁵ Il logit è una funzione che si applica a valori probabili compresi nell'intervallo tra 0 e 1. Viene definito come: $\text{logit}(p) = \ln \left(\frac{p}{1-p} \right) = \ln(p) - \ln(1-p)$, in cui \ln è il logaritmo naturale e $\left(\frac{p}{1-p} \right)$ è l'*odds*, ossia il rapporto tra la probabilità p di un evento e la probabilità $(1-p)$ dell'evento complementare. Per logaritmo naturale o logaritmo di Nepero si intende il logaritmo in base e . Il

segue:

Formula n. 3.41:
$$\text{logit}(P_j(\theta_i)) = \ln\left(\frac{P_j(\theta_i)}{1 - P_j(\theta_i)}\right) = \theta_i - b_j$$

In questo modo la misura non è proporzionale alla probabilità di risposta ma alla corrispondente trasformazione in logit. Sia la difficoltà che l'abilità vengono misurate su una stessa scala che utilizza il logit, l'unità di misura basata sul logaritmo dell'*odds*. In questo modo diventa facile comparare le due variabili, se il livello di abilità e di difficoltà hanno la stessa misura, la differenza tra il primo e il secondo fornisce la probabilità di rispondere in maniera corretta in logit. Se, ad esempio, un candidato con un'abilità pari a 1 logit affronta un *item* di difficoltà pari a 1 logit, la differenza è pari a 0, dunque ha una probabilità del 50%; se il candidato ha un'abilità di 1 logit e un *item* di difficoltà pari a 2, l'abilità è inferiore alla difficoltà di 1 logit, dunque la probabilità di rispondere si riduce.

Dal punto di vista grafico la ICC si può rappresentare mediante un diagramma cartesiano; si tratta generalmente di una funzione crescente: al crescere del livello di abilità latente (θ), aumenta la probabilità ($P_j(\theta_i)$) di rispondere all'*item* in modo corretto, come rappresentato dalla figura in basso:

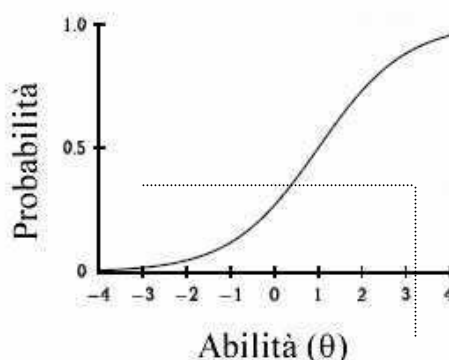


Figura n. 3.12 - Rappresentazione della *Item Characteristic Curve* (ICC)

logit ha come funzione inversa: $p = \left(\frac{e^{\text{logit}}}{1 + e^{\text{logit}}}\right)$.

Il piano cartesiano della figura n. 3.12 rappresenta, sull'asse verticale, la probabilità ($P_j(\theta_i)$) di rispondere correttamente all'*item* (che può variare all'interno di un intervallo compreso tra 0 e 1) e, sull'asse orizzontale, l'abilità latente (θ) stimata (su di una scala in cui la media è posta sullo 0). La figura n. 3.12 mostra dunque l'andamento della ICC al crescere del valore di (θ). Si può notare visivamente che la probabilità ($P_j(\theta_i)$) aumenta al crescere dell'abilità ($\theta_i - b_j$). Però ($P_j(\theta_i)$) aumenta in modo non lineare: se l'abilità del candidato è uguale alla difficoltà dell'*item*, si ha il 50% di probabilità di una risposta corretta (ossia un valore uguale a 0,5); se la difficoltà dell'*item* è maggiore del livello di abilità posseduto dal candidato, la probabilità di rispondere correttamente è al di sotto di 0,5; se, al contrario, la difficoltà dell'*item* è minore del livello di abilità del candidato, la probabilità di rispondere correttamente è al di sopra di 0,5. In sintesi, più è alto il parametro della difficoltà, più è alto il livello di abilità richiesto per rispondere in modo corretto. Di conseguenza, la ICC di un *item* difficile è generalmente posizionata nella parte destra della scala delle abilità, quella di un *item* facile nella parte sinistra della scala.

Quando la scala delle abilità è standardizzata (per cui la media è pari a 0 e la deviazione standard è pari a 1) il valore di b_j (difficoltà dell'*item*) varia tra -2,00 e +2,00. I valori vicino al -2,00 indicano *item* piuttosto facili, invece i valori vicini al +2,00 si riferiscono agli *item* più difficili.

Supponendo, ad esempio, di avere un'abilità con una media pari a 0 e una deviazione standard pari a 1, e tre candidati (A, B, C) con abilità diverse, le probabilità di risposta dei tre candidati, illustrate nella figura sottostante (fig. n. 3.13), sono prevedibilmente diverse.

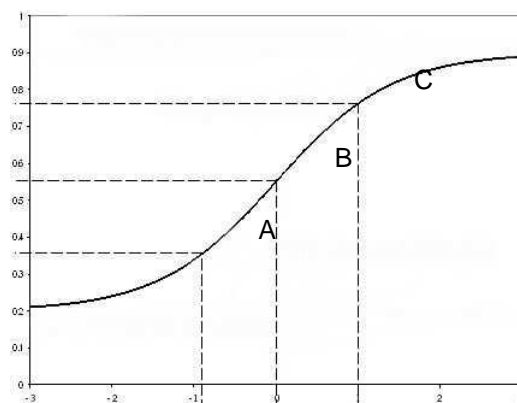


Figura n. 3.13 - Un esempio di rappresentazione della *Item Characteristic Curve*

Il candidato A, con un'abilità pari a -0,9 ha una probabilità di rispondere in modo corretto pari a circa 0,35; il candidato B, con un'abilità pari a 0 ha una probabilità di rispondere in modo corretto pari a circa 0,55; il candidato C, con un'abilità pari a +1,00 ha una probabilità di rispondere in modo corretto pari a circa 0,75.

È chiaro che, se più candidati hanno la stessa abilità, hanno anche la stessa probabilità di rispondere correttamente all'*item*.

In questo modello a un parametro, come già accennato, non viene mai considerato il parametro di *guessing*, per cui il valore più basso della ICC è 0, dunque i candidati con un livello di abilità che tende a valori bassi hanno una probabilità di rispondere in modo corretto all'*item* che tende a zero.

Il punteggio grezzo per ogni risposta corretta, pari a 1 per ogni *item* (modello dicotomico), mediante il modello di Rasch, viene trasformato nella probabilità attesa (quindi in un valore compreso nell'intervallo continuo tra 0 e 1) di osservare una risposta pari a 1. La trasformazione in logit (dunque in un valore compreso nell'intero asse reale) si rivela necessaria per evitare che a candidati con estremi differenti livelli di abilità venga assegnata la medesima probabilità.

Il modello di Rasch sostanzialmente stima la difficoltà di ciascun *item* e

l'abilità di ciascun candidato in logit. Ciò consente di creare una graduatoria di difficoltà degli *item* e una graduatoria di difficoltà dei candidati, confrontabili tra loro e al loro interno. In altri termini, è possibile stabilire quanto un *item* è più o meno difficile rispetto a un altro e quanto un individuo è più o meno abile rispetto a un altro; altresì è possibile confrontare l'abilità di un individuo con la difficoltà di un *item* in modo da prevedere la probabilità di scelta della risposta.

In un piano cartesiano è possibile rappresentare più di una ICC (fig. n. 3.14):

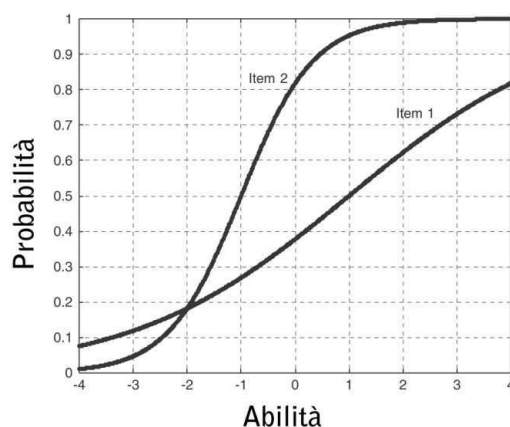


Figura n. 3.14 - Esempio di ICC di due *item*

La rappresentazione mostra due ICC, ogni curva è caratteristica del proprio *item* (1, 2).

La quantità di informazioni che un *item* fornisce per la stima dell'abilità dei candidati è detta **Item Information Function** (IIF) e viene rappresentata dal simbolo " $I_i(\theta)$ ". La quantità di informazioni fornite da un *item* è uguale alla probabilità di una risposta corretta moltiplicata per la probabilità di una risposta errata:

Formula n. 3.42:
$$I_i(\theta) = p_i(\theta) q_i(\theta)$$

Uno stesso *item* fornisce informazioni diverse in base al differente livello di abilità dei vari candidati, dunque $I_i(\theta)$ può essere rappresentata da una curva

simile a quella illustrata di seguito:

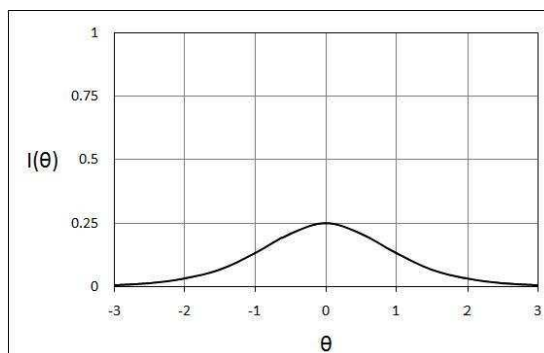


Figura n. 3.15 - Rappresentazione della *Item Information Function* (IIF)

La IIF fornisce un'idea sul punto dell'asse dell'abilità in cui l'*item* è maggiormente informativo: un *item*, quindi, è molto informativo solo per uno specifico livello. Nel modello di Rasch il punto in cui l'informazione dell'*item* è massima corrisponde al livello di difficoltà dell'*item* stesso. Più IIF in uno stesso piano cartesiano mostrano l'*item* che fornisce più informazioni e soprattutto a quale livello dell'abilità funziona meglio.

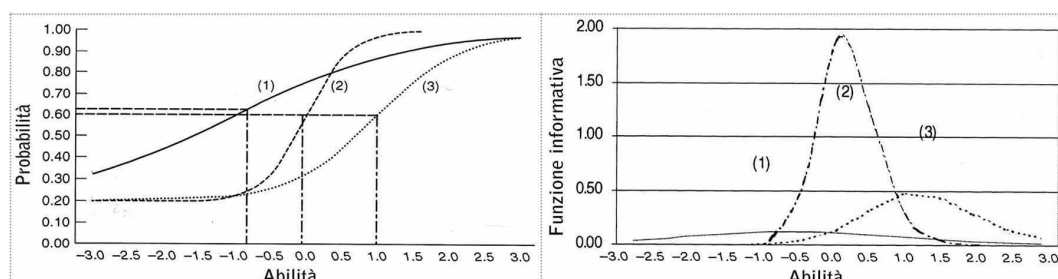


Figura n. 3.16 – Esempio di ICC e IIF (Bachman, 2004:141-146)

La figura n. 3.16 mostra un esempio in cui sono rappresentati tre *item*: il secondo *item* appare il più informativo (fig. a destra), inoltre il punto in cui l'informazione dell'*item* è massima corrisponde al livello di abilità più elevato, e dunque al livello di difficoltà dell'*item* stesso.

Se si tengono in considerazione tutte le IIF di ogni *item* di uno stesso test, è possibile calcolare la funzione informativa del test nella sua globalità, ovvero la **Test Information Function** (TIF), rappresentata dal simbolo " $I(\theta)$ " e calcolata

mediante la somma delle IIF:

Formula n. 3.43: $I(\theta) = \sum I_i(\theta)$

La TIF è strettamente connessa al numero degli *item* contenuti nel test, quindi quanto più il numero degli *item* aumenta tanto più accresce l'informazione del test e, dunque, la precisione con cui l'abilità viene stimata.

Una caratteristica della TIF consiste nel contributo apportato da ogni *item* in maniera indipendente dagli altri, che concorre alla stima del singolo contributo sull'informazione totale (diversamente nella CTT l'effetto di un particolare *item* sull'informazione totale dipende dal grado di correlazione con gli altri *item*).

Tra la CTT e l'IRT vi sono molte differenze, indicate in modo implicito in queste pagine e qui di seguito sintetizzate. Innanzitutto la CTT presuppone una relazione di tipo lineare, laddove l'IRT ipotizza una relazione di tipo non lineare. In secondo luogo nella CTT la difficoltà di un *item* è data dalla proporzione dei candidati che hanno risposto correttamente all'*item* stesso, invece nel modello di Rasch, in cui il solo parametro considerato è il livello di difficoltà dell'*item*, questo corrisponde al punteggio, nella scala di abilità, in cui la probabilità di rispondere in maniera corretta a un *item* è uguale alla probabilità di rispondere in maniera errata. Infine la CTT si concentra sull'informazione fornita dal test nella sua totalità, viceversa l'IRT si sofferma sull'informazione fornita da ogni singolo *item*. Alla luce di ciò, appare evidente che la IRT offre maggiori vantaggi rispetto alla CCT e si presenta come un potente strumento di misurazione.

Ciò nonostante, la difficoltà (dal punto di vista statistico) nell'utilizzare i complessi modelli dell'IRT induce talvolta all'adozione della CTT. Persino il modello di Rasch, il più semplice tra i modelli IRT, è utilizzato da esperti statistici. Un'altra problematicità del modello di Rasch è l'attribuzione di stime relative al

livello di abilità dei candidati e al livello di difficoltà agli *item*, non sempre empiricamente relazionabili in modo simmetrico. Un ulteriore punto critico attribuibile al modello di Rasch è l'attenzione che pone unicamente sul parametro di difficoltà dell'*item*, ignorando il parametro di discriminazione e il parametro di *guessing*. A dispetto di tali spinosità e criticità dell'IRT, non si può negare che tale analisi fornisce risultati sicuramente più precisi.

3.3.3 Calcolo dell'affidabilità

Un test è affidabile quando misura il livello di abilità del candidato mantenendo costante il risultato ottenuto anche se la procedura di misurazione viene replicata nel tempo. Più un test è attendibile, più garanzie di coerenza e costanza offre sui risultati.

Poiché un test non può definirsi totalmente indipendente da possibili fonti di errore, nel momento in cui si conduce un'analisi statistica sui dati prodotti da una prova linguistica e si misura la stabilità delle distribuzioni dei punteggi ottenuti, si deve tenere grande considerazione dell'errore di misurazione, ossia dell'influenza che le fonti di incoerenza (descritte nel paragrafo 2.8) hanno sulla variabilità dei dati.

Esistono modelli di misurazione diversi, legati alle teorie sia classica che di risposta degli *item* che della generalizzabilità, le quali permettono di stimare gli effetti di queste fonti di incoerenza che condizionano i punteggi di un test e l'affidabilità del test stesso.

Nella CTT il **coefficiente di affidabilità**, rappresentato dal simbolo " r_{xx} ", indica il rapporto tra la varianza vera (σ_v^2) e la varianza totale (σ^2), ossia la

somma della varianza vera (σ^2_v) e della varianza dovuta all'errore casuale (σ^2_e), dei punteggi dei test:

Formula n. 3.44:
$$r_{xx} = \frac{\sigma^2_v}{\sigma^2_v + \sigma^2_e}$$

L'affidabilità permette di stimare qual è la proporzione di varianza vera e quale la proporzione di varianza erratica rapportata alla varianza totale, dunque il grado di influenza che i punteggi veri e quelli dovuti a errore hanno sui punteggi totali. Maggiore è la proporzione di varianza vera, minore è la varianza dovuta a errore, e dunque maggiore è il coefficiente di affidabilità.

I dati di un test forniscono informazioni sull'abilità del candidato, per cui sembra chiaro che la varianza dei punteggi reali risulti maggiore della varianza erratica. Graficamente pertanto tale proporzione può essere rappresentata come segue:

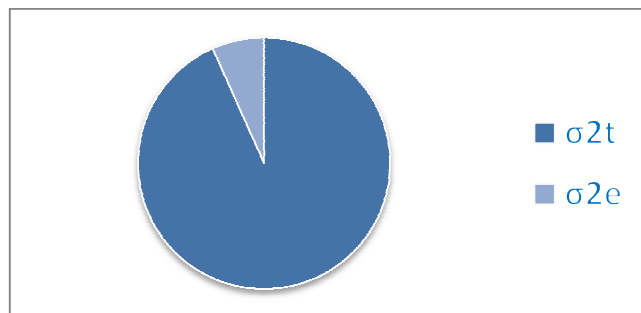


Figura n. 3.17 - Varianza nel modello CTT (Bachman, 2004:158)

Nel modello della CTT la **varianza totale** (σ^2) è data dalla somma della varianza dei punteggi reali (σ^2_v) e della varianza dovuta a errori di misurazione (σ^2_e), secondo la formula che segue:

Formula n. 3.45:
$$\sigma^2 = \sigma^2_v + \sigma^2_e$$

La CTT offre vari metodi per stimare l'affidabilità di un test in base al tipo di fonte di incoerenza: l'incoerenza interna tra gli *item* di uno stesso test, l'instabilità durante il tempo, l'incoerenza tra le varie forme di un test e

l'incoerenza tra i valutatori. Il metodo della CTT fornisce due approcci per la valutazione dell'affidabilità per ognuno delle quattro fonti di errore: il calcolo della correlazione e il calcolo del **coefficiente alpha** (α), detto anche "alpha di Cronbach".

Per valutare la coerenza interna (*inter-item consistency*) tra gli *item* di uno stesso test è possibile seguire il metodo "split-half" oppure il metodo che prende in considerazione la varianza.

Se ci si avvale del primo metodo spesso si decide di somministrare il test dividendolo in due parti omogenee tra loro e in seguito calcolarne l'affidabilità correlando i risultati delle due metà. Un test si può dividere a metà in maniera casuale oppure attraverso la selezione di *item* pari da una parte e di *item* dispari dall'altra. È necessario però controllare che gli *item* siano indipendenti e che misurino gli stessi aspetti della stessa abilità. Si dovrebbero ottenere due metà abbastanza equivalenti, due *sub-test* con media e varianza abbastanza simili, come se fossero due test paralleli. Le due distribuzioni di punteggi ottenute dalle due metà vengono in seguito correlate. Al momento della stima dell'affidabilità è possibile affidarsi, per calcolare il coefficiente di correlazione, alla formula di Spearman-Brown oppure alla formula di Guttman.

Il coefficiente di correlazione fra le due distribuzioni rappresenta però il grado di affidabilità solo di metà test. Per l'affidabilità dell'intero test si utilizza la formula di Spearman-Brown che aiuta a correggere il problema della lunghezza del test, la quale incide sull'affidabilità del test stesso (poiché un test è tanto più affidabile quanto più è lungo in termini di numero di *item*, dividendolo in due metà, l'affidabilità tende a diminuire rispetto all'affidabilità totale). La stima del **coefficiente di affidabilità** dell'intero test si definisce con la formula che segue:

Formula n. 3.46:
$$r_{xx} = \frac{2r_{AB}}{1 + r_{AB}}$$

in cui r_{AB} indica il coefficiente di correlazione tra le due parti (A e B) in cui il test è stato suddiviso.

La formula di Guttman fornisce l'indice di affidabilità dell'intero test e si rappresenta mediante la formula:

Formula n. 3.47:
$$r_{xx} = 2 \left(1 - \frac{\sigma^2_A + \sigma^2_B}{\sigma^2_x} \right)$$

in cui σ^2_A e σ^2_B sono le varianze delle due metà del test e σ^2_x è la varianza dell'intero test.

Per sapere quanto aumenta l'affidabilità in proporzione al numero di *item* è possibile applicare una particolare formula di Spearman e Brown, detta "formula profetica" (*prophecy formula*):

Formula n. 3.48:
$$r_{xx} = \frac{kr_{AB}}{1 + (K-1)r_{AB}}$$

in cui K indica il numero degli *item* e r_{AB} indica il coefficiente di correlazione tra le due parti (A e B) del test.

La formula profetica mostra come il livello di affidabilità dipenda anche dal numero di *item*, dunque dalla lunghezza del test. Per applicare tale formula si assume che gli *item* aggiunti abbiano le stesse caratteristiche di affidabilità degli *item* iniziali. Tale formula consente di calcolare le variazioni del coefficiente di affidabilità di un test in funzione dell'aumento (o della diminuzione) del numero di *item*.

Se si possiede un livello di affidabilità desiderato, e si vuol calcolare di quanto allungare il test (numero di *item* necessari per un test affidabile) per ottenere tale livello, si può applicare la seguente formula:

Formula n. 3.49:
$$k = \frac{r_{tt}(1-r_{xx})}{r_{xx}(1-r_{tt})}$$

in cui r_{tt} è il livello di affidabilità desiderato e r_{xx} è l'affidabilità stimata per il test iniziale. Se si moltiplica poi il numero degli *item* iniziali per il valore di K ottenuto, si avrà il numero degli *item* necessari per il nuovo test.

Se, per la valutazione della coerenza interna tra gli *item* di uno stesso test, ci si avvale del metodo della varianza, è necessario distinguere test con domande a credito parziale per cui si adotta l'*alpha* di Cronbach, da test di tipo dicotomico (domande di tipo giusto/sbagliato, le cui risposte sono valutate con 1 o 0) per cui si utilizzano le formule elaborate da Kuder e Richardson (conosciute come KR-20 e KR-21).

La formula approntata da Cronbach è impiegata per valutare il grado di affidabilità di un test a partire dai singoli *item*, ossia consente di valutare l'apporto di ogni singolo *item* al grado di affidabilità, ed è espressa dall'indice *alpha*:

Formula n. 3.50:
$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum \sigma_i^2}{\sigma_x^2} \right)$$

in cui K è il numero degli *item* o delle parti in cui il test è stato suddiviso, $\sum \sigma_i^2$ è la somma delle varianze di ogni *item*, σ_x^2 è la varianza totale.

Le formule sviluppate da Kuder e Richardson sono casi speciali di coefficiente *alpha* per test di tipo dicotomico. La formula KR-20 è:

Formula n. 3.51:
$$r_{xx} = \frac{k}{k-1} \left(1 - \frac{\sum p_i q_i}{\sigma_x^2} \right)$$

in cui K è il numero degli *item*, p_i indica la proporzione di risposte esatte all'*item*, q_i indica la proporzione di risposte errate all'*item*, $\sum p_i q_i$ è la somma delle varianze degli *item*, σ_x^2 è la varianza totale dei punteggi nel test.

Una formula più semplice della precedente, nonostante equivalente, è la

KR-21, in cui non è necessario calcolare la varianza di tutti gli *item*:

$$\text{Formula n. 3.52:} \quad r_{xx} = \frac{k}{k-1} \left(1 - \frac{\sum \bar{p} \bar{q}}{\sigma^2_x} \right)$$

in cui K è il numero degli *item*, \bar{p} indica la media dei p_i , \bar{q} indica la media dei q_i .

Una ulteriore semplificazione dei calcoli è possibile con la seguente versione della formula KR-21:

$$\text{Formula n. 3.53:} \quad r_{xx} = \frac{k \sigma^2 - (k - \bar{x})}{(k-1)\sigma^2}$$

in cui K è il numero degli *item*, \bar{x} è la media dei punteggi e σ^2 indica la varianza dei punteggi nel test. L'utilizzo di tale formula consente di calcolare solo la media e la varianza dei punteggi totali del test.

Una seconda possibile fonte di errore nella valutazione dell'affidabilità è l'instabilità durante il tempo. Per valutare l'affidabilità di un test in questo caso ci si avvale del metodo che richiede due somministrazioni dello stesso test (detto "test-retest"), per cui è possibile somministrare lo stesso test due volte allo stesso gruppo di candidati, con un breve lasso di tempo tra una somministrazione e l'altra, per verificare che i risultati restino stabili nel tempo.

L'affidabilità è data dalla correlazione tra le due serie di punteggi ottenuti. Più è alto il coefficiente di correlazione, più il test e i risultati ottenuti risulteranno affidabili.

La durata del lasso di tempo che intercorre tra una somministrazione e un'altra è fondamentale ma sostanzialmente difficile da stabilire. Se il lasso di tempo è troppo breve il coefficiente di correlazione può essere inficiato dal ricordo che i candidati hanno delle risposte data agli *item* nella prima somministrazione. Se il lasso di tempo, invece, è troppo lungo il coefficiente di correlazione apparire inficiato dal cambiamento delle competenze dei candidati

(sia positivamente che negativamente). Secondo gli psicometristi l'intervallo di tempo adatto non deve essere inferiore a un mese e non deve andare oltre i sei mesi circa.

Va sottolineato però che la variabilità dell'intervallo dipende anche dalla tipologia dei candidati presi in esame, dall'abilità misurata e dal tipo di processo di apprendimento a cui i candidati si sottopongono nel periodo di tempo che intercorre tra le due somministrazioni. Una volta attuate le due somministrazioni e ottenute le due distribuzioni di punteggi, è possibile stimare la stabilità dei risultati attraverso il calcolo del coefficiente di correlazione oppure attraverso il calcolo del coefficiente *alpha*:

Formula n. 3.54:
$$r_{xx} = \left(\frac{\sigma^2_{x1} + \sigma^2_{x2}}{\sigma^2_{x1+x2}} \right)$$

in cui σ^2_{x1} indica la varianza dei punteggi della prima somministrazione, σ^2_{x2} indica la varianza dei punteggi della seconda somministrazione e σ^2_{x1+x2} è l'insieme della varianza dei punteggi della prima e seconda somministrazione.

La terza fonte di errore possibile è l'incoerenza tra le forme parallele di un test. Per poter stimare l'affidabilità di due test equipollenti occorre sviluppare due forme di test equivalenti dal punto di vista del contenuto e dell'abilità che si intende misurare, e in seguito somministrare le due forme allo stesso gruppo di candidati. È manifesto che le due forme non possono essere somministrate contemporaneamente ma, al tempo stesso, non è possibile far trascorrere un breve lasso di tempo poiché si confonderebbe l'incoerenza dovuta al tempo trascorso (ricordo della prima somministrazione) con l'incoerenza causata dalle forme dei due test.

Per ovviare a questo problema è possibile servirsi di un programma di contrappeso, ossia si possono somministrare due forme parallele di un test a due

gruppi di candidati pressoché equivalenti, facendo in modo che il primo gruppo si sottoponga al test A in prima istanza e al test B in seconda; il secondo gruppo al test B in prima istanza e al test A in seconda. Successivamente sarà possibile misurare l'affidabilità dei risultati mediante il calcolo del coefficiente di correlazione o del coefficiente *alpha* attraverso la formula n. 3.54 precedentemente descritta.

L'ultima fonte di incoerenza che può compromettere l'affidabilità di un test è costituita dai valutatori. Per la stima sia della *intra-rater* che della *inter-rater reliability* (si rimanda, per la definizione dei due concetti, al paragrafo 2.8.3) è opportuno sottoporre i dati di ciascun test a più valutazioni. Nel caso della *intra-rater reliability* si rende necessario far valutare le prove da uno stesso valutatore più di una volta a distanza di breve tempo. Nel caso della *inter-rater reliability* è indispensabile ottenere valutazioni, per lo stesso test, da più valutatori. Una volta ottenute le due distribuzioni di punteggi è possibile calcolare l'affidabilità attraverso l'indice di correlazione e il coefficiente *alpha*. Nel caso in cui si hanno più valutazioni per lo stesso candidato si può scegliere di raggruppare le valutazioni attraverso il calcolo della media o attraverso la mera somma delle valutazioni. In tal caso è più opportuno calcolare il coefficiente *alpha* impiegando la formula che segue:

Formula n. 3.55:
$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum \sigma_{ri}^2}{\sigma_x^2} \right)$$

in cui K è il numero delle valutazioni, $\sum \sigma_{ri}^2$ indica la somma delle varianze delle valutazioni individuali, σ_x^2 è la varianza di tutte le valutazioni raggruppate (attraverso il calcolo della somma o della media).

Le formule riguardanti il calcolo dell'affidabilità introdotte finora si riferiscono alla misurazione di punteggi relativi a un gruppo. Nessuna fornisce

informazioni sui punteggi individuali, per le quali è necessario impiegare la formula per il calcolo dell'**errore standard di misura** (SEM):

$$\text{Formula n. 3.56:} \quad \mathbf{SEM = DS \sqrt{1 - r_{xx}}}$$

in cui DS indica la deviazione standard e r_{xx} è il coefficiente di affidabilità.

Si tratta di stimare la deviazione standard dei punteggi osservati di un candidato dopo varie somministrazioni di uno stesso test (o di forme parallele di un test) nelle medesime condizioni, ossia, diversamente dalla DS che descrive la variabilità dei risultati, il SEM descrive il grado di incertezza nella stima dei risultati, il margine di errore connesso alla stima stessa.

È possibile utilizzare il calcolo del SEM per interpretare l'effetto che l'errore ha sul risultato individuale. Se si considera che il punteggio grezzo è il reale punteggio del candidato, è possibile calcolare un **intervallo di confidenza** o fedeltà, cioè l'intervallo compreso tra i limiti (detti "limiti fiduciali") inferiore e superiore, entro i quali si ha una certa probabilità che cada il vero valore del parametro. L'operazione si può eseguire con la formula che segue:

$$\text{Formula n. 3.57:} \quad \mathbf{IC_p = X \pm z_p (SEM)}$$

in cui p indica il livello di probabilità o confidenza per cui si calcola l'intervallo di confidenza, X indica il punteggio grezzo, z_p è il numero dei punti Z associati a p (ossia il numero di deviazioni standard necessario per ottenere l'intervallo per un certo livello di confidenza) e il SEM è l'errore standard di misura.

I valori più comuni per indicare il livello di probabilità sono 0,68, 0,95 e 0,99. I punti Z associati a tali livelli di probabilità sono rispettivamente 1,00, 1,96 e 2,58. Il livello di probabilità più appropriato dipende dal grado di affidabilità che si vuol dare alla stima del punteggio.

Se, ad esempio, si ha un coefficiente di affidabilità pari a $r_{xx} = 0,75$, una

deviazione standard uguale a $DS = 4,5$, il SEM si calcola come: $SEM = 4,5 \sqrt{1 - 0,75} = 2,25$. Per calcolare l'intervallo di confidenza con una probabilità di 0,95, si ha $IC_{0,95} = X \pm 1,96 (2,25) = X \pm 4,4$. Se si presume che il punteggio grezzo X è uguale a 30, si ottiene $IC_{0,95} = 30 \pm 4 = 26 - 34$. L'intervallo di confidenza entro cui cade il valore sta tra il limite inferiore di 26 e il limite superiore di 34.

In conclusione, la CTT offre diversi metodi per stimare l'affidabilità di un test, la scelta di uno piuttosto che di un altro dipende dall'identificazione della fonte di errore di misurazione.

La CTT presenta però diversi limiti, primo fra tutti ipotizza sempre che l'errore sia casuale e non fa distinzione tra errore casuale ed errori sistematici (quali le caratteristiche personali del candidato, le caratteristiche del metodo e della somministrazione di un test).

Secondariamente permette il calcolo dell'affidabilità e del margine di errore (SEM) per tutti i livelli di abilità in modo uniforme, invece i punteggi non sono equamente affidabili (diversamente la IRT offre la possibilità di stimare diversi errori per i diversi livelli di abilità).

Infine, per calcolare l'affidabilità, si può considerare una fonte di errore alla volta. In realtà, invece, in una situazione valutativa possono coesistere o manifestarsi più fonti di errore contemporaneamente. Si potrebbero esaminare singolarmente, tuttavia si perderebbe l'interazione tra le fonti.

A ciò può porre rimedio la "teoria della generalizzabilità" (*G-theory*), la quale rende possibile la stima dell'affidabilità prendendo in esame gli effetti di più di una fonte di errore simultaneamente, permette infatti di identificare l'errore casuale e altresì l'errore sistematico, e di eliminare le fonti di questo tipo di

errore. La teoria della generalizzabilità permette di stimare gli effetti di diversi aspetti, detti “*facets*”, appartenenti alla procedura di misurazione. Un *facet* di misurazione è una fonte di variazione associata a un aspetto della procedura di misurazione e interpretata come una potenziale sorgente di errore di misurazione. I *facet* sono insiemi di condizioni simili, casuali e parallele, per cui se un test è costituito da un insieme di *item*, questi ultimi costituiranno un *facet*, dunque una fonte di probabile errore.

La teoria della generalizzabilità assume l'esistenza di un universo ipotetico di misure (gli *item*) dal quale viene estratto in modo casuale un campione per la somministrazione. Per poter inferire sull'intero universo ipotetico è necessario che il campione sia rappresentativo di tutto l'universo. Pertanto si rende indispensabile definire l'universo su cui si vuole generalizzare. Coloro che si occupano di sviluppare il test specificano l'universo delle misure ammissibili quali indicatori dell'abilità da misurare (tali misure devono soddisfare le specificazioni predisposte per il dato test); coloro che sono responsabili delle decisioni da prendere sulla base dei risultati e coloro che si occupano di inferire sulla *performance* del candidato specificano l'universo dei compiti, attraverso i quali si potranno effettuare le generalizzazioni sulle competenze linguistiche dei candidati. In una situazione ideale coloro che si occupano di sviluppare il test svolgono il loro lavoro sulla base dell'universo dei compiti predisposto da coloro che si occupano di inferire sulla *performance* del candidato.

Quando si progetta un test, in genere, si hanno già in mente i destinatari, pertanto è possibile specificare anche l'universo degli individui, detti “oggetto della misurazione”, che si sottoporranno al test. Le variazioni dei punteggi sono dunque attribuibili al diverso livello di abilità di ogni individuo. Ogni punteggio osservato è un campione di *performance* dell'universo di generalizzazione di un

dato individuo, ossia quello che nella CTT è indicato come punteggio reale e nella G-theory è l'“*universe score*” (se un candidato viene sottoposto alla valutazione di più valutatori, ad esempio, il suo *universe score* è dato dalla media di tutti i punteggi osservati). Non essendo possibile osservare l'intero universo, la sua misura può essere solo stimata attraverso i dati osservabili.

Ma quanto è attendibile la stima dell'*universe score* operata mediante il punteggio osservato?

La stima dell'affidabilità deve affrontare due problematiche: la stima del punteggio dell'universo e la valutazione della generalizzabilità di tale stima (operata mediante una particolare applicazione dell'analisi della varianza). Insieme con essa va valutato anche il margine di errore presente in tale stima. Date le possibili variazioni nelle misurazioni, per ottenere misure stabili e coerenti, è necessario disporre di misure multiple, individuate a partire da una popolazione teorica di misure possibili, utilizzabili per stimare la vera misura dell'abilità, tenendo sotto controllo gli errori casuali.

Utilizzando un campione di misure multiple, il primo obiettivo è quello di stimare il punteggio che si sarebbe ottenuto se fosse stato impiegato tutto l'universo di misure.

La possibilità che su casi diversi vengano utilizzate misure multiple differenti conduce alla distinzione tra diversi disegni. Il cosiddetto “*crossed design*” è attuabile quando lo stesso campione casuale di misure multiple estratto dall'universo viene applicato a tutti i casi. Se, ad esempio, tutti i candidati rispondono a tutti gli *item*, la varianza (σ^2) può essere determinata dai candidati (*c*), dagli *item* (*i*) e dall'interazione tra i candidati e gli *item* (*ci*). La varianza totale è dunque data da:

Formula n. 3.58:
$$\sigma^2 = \sigma^2_c + \sigma^2_i + \sigma^2_{ci,e}$$

in cui σ^2_c è la varianza determinata dai candidati, σ^2_i indica la varianza determinata dagli *item* e $\sigma^2_{ci,e}$ è la varianza determinata dall'interazione tra i candidati e gli *item* e da altri fattori che possono incidere sul punteggio generando errore, ma che non possono essere distinti dall'interazione tra i candidati e gli *item* (*ci*). Se vi sono 50 *item* e 50 candidati, e ogni candidato risponde ai 50 *item*, si ottengono 50 punteggi per ogni individuo, in modo da poter stimare la varianza di ogni individuo, si ottengono inoltre tanti punteggi per ogni *item* per quanti sono i candidati, in modo da poter stimare la *varianza* per ogni *item*. Avendo un punteggio totale per ogni candidato, ottenuto dalla combinazione degli *item*, non è possibile distinguere la varianza determinata dall'interazione tra candidato e *item* dall'errore casuale.

Il cosiddetto “*nested design*” è attuabile quando un diverso gruppo di misure multiple, casualmente estratto dall'universo, viene applicato a ciascun caso. Se, ad esempio, si somministra un test sulla comprensione scritta costituito da domande associate a determinati passaggi, tali *item* sono utilizzabili solo con quel dato testo, pertanto gli *item* sono legati a quel particolare test. In questo caso la varianza totale (σ^2) può essere determinata dai candidati (*c*), dai testi (*t*), dall'interazione tra i testi e gli *item* (*i:t*), dall'interazione tra i candidati e i testi (*pt*), e dall'interazione tra i candidati e gli *item*, che includono a loro volta i testi ad essi legati, (*pi:t*):

Formula n. 3.59:
$$\sigma^2 = \sigma^2_c + \sigma^2_t + \sigma^2_{i:t} + \sigma^2_{pt} + \sigma^2_{pi:t,e}$$

L'affidabilità della misura finale dipende così dal numero di misure multiple e dalla relazione tra le misure multiple e l'abilità generale da misurare. Nella teoria della generalizzabilità, contrariamente alla teoria classica che

ammette un solo coefficiente di affidabilità, i punteggi possono avere vari coefficienti di generalizzabilità (affidabilità), a seconda dei fattori che, influenzando il procedimento di misurazione, vengono presi in considerazione. In ogni caso la definizione di **generalizzabilità** (ρ^2) è, come per la CTT (formula n. 3.44), data dal rapporto tra la varianza del punteggio dell'intero universo σ_p^2 e la varianza totale σ^2 (ossia la somma della σ_p^2 e della varianza dell'errore σ_e^2):

Formula n. 3.60:
$$\rho^2 = \frac{\sigma_p^2}{\sigma_p^2 + \sigma_e^2}$$

La G-theory permette di stimare gli effetti di molteplici sorgenti di errore allo stesso tempo. Tuttavia, come la CTT, ipotizza che l'errore di misurazione sia lo stesso per tutti i livelli di abilità.

La IRT, come già discusso, permette invece la stima dell'errore per ogni livello di abilità e inoltre, diversamente dalla CCT e *G-theory*, l'affidabilità stimata non dipende dal particolare campione di candidati presi come riferimento; nell'IRT il punteggio ottenuto da un soggetto è indipendente dalla specifica prova somministratagli e le caratteristiche di un *item* sono indipendenti rispetto al particolare gruppo di individui cui l'*item* è destinato. Il valore di TIF ($I(\theta)$) (calcolato mediante la formula n. 3.43) può essere utilizzato per calcolare l'**errore standard**:

Formula n. 3.61:
$$SE(\hat{\theta}) = \frac{1}{\sqrt{I(\theta)}}$$

in cui $SE(\hat{\theta})$ indica l'errore standard per ogni livello di abilità (θ).

La differenza rispetto al calcolo del SEM della CTT consiste proprio nell'irregolarità e nella discontinuità dell'errore in tutti i livelli di abilità.

Le procedure di misurazione dell'affidabilità descritte finora sono più adeguate per le prove di tipo normativo, giacché i risultati sono interpretati sulla

base di un confronto con i risultati ottenuti da un gruppo di individui, che rappresenta il modello assunto come riferimento al quale uniformarsi. Al contrario nei test di tipo criteriale i risultati sono interpretati sulla base di criteri di riferimento prestabiliti, indipendentemente dal particolare gruppo di individui che si è sottoposto al test. Di conseguenza per quest'ultima tipologia di prova occorre impiegare procedure diverse per la stima dell'affidabilità.

Il **coefficiente di affidabilità**, rappresentato dalla lettera greca "phi" (Φ), è calcolabile mediante la formula KR-20 per le risposte dicotomiche, oppure mediante il coefficiente *alpha*, perché è una formula più generica:

Formula n. 3.62:
$$\Phi = \frac{\frac{ns_p^2(\alpha)}{n-1}}{\frac{ns_p^2(\alpha)}{n-1} + \frac{\bar{X}_p(1-\bar{X}_p) - s_p^2}{k-1}}$$

in cui n è il numero dei candidati, k indica il numero di *item* (nel caso di *item* dicotomici) o il punteggio più alto possibile (nel caso di *item* a credito parziale), \bar{X}_p è la media dei punteggi, s_p^2 è la varianza dei punteggi, α indica il coefficiente *alpha*. La formula permette di stimare il valore di affidabilità di un punteggio, quale indicatore del livello di abilità di un candidato rispetto al criterio prestabilito.

Come per i test normativi, anche per i test criteriali è possibile calcolare l'errore standard (SEM):

Formula n. 3.63:
$$SEM = \sqrt{\frac{\bar{X}_p(1-\bar{X}_p) - s_p^2}{k-1}}$$

Per calcolare l'intervallo confidenziale entro cui si ha una certa probabilità che cada il vero punteggio si può impiegare la formula n. 3.57 sopraesposta (i valori per i livelli di probabilità più comuni restano uguali e, di conseguenza, anche i punti Z). È necessario ricordare però che il SEM della formula n. 3.63 è calcolato sulla proporzione dei candidati che ha risposto in modo corretto,

pertanto, prima di calcolare l'intervallo confidenziale, è necessario riportare il valore al punteggio grezzo, moltiplicando il SEM per il numero degli *item* del test.

È possibile anche calcolare l'errore standard esclusivamente per un individuo mediante la seguente formula:

Formula n. 3.64:
$$SE_{\text{means}(X_i)} = \sqrt{\frac{X_i(k-X_i)}{k-1}}$$

in cui k indica il numero degli *item* (nel caso di *item* dicotomici) o il punteggio più alto possibile (nel caso di *item* a credito parziale), X_i indica il punteggio di un individuo.

3.3.4 Calcolo della validità

La ragione per cui si somministra un test e si analizzano i risultati consiste nella possibilità di inferire sulle conoscenze di uno o più candidati che vanno al di là del dominio di conoscenze linguistiche presenti nel test. Le inferenze possono definirsi corrette se i risultati sono validi, ossia se sono reali indicatori di ciò che si vuol misurare.

Per poter affermare che i risultati riflettono reali competenze occorre darne evidenza empirica. Si tratta di congiungere il concetto teorico e astratto di validità (esposto nel paragrafo 2.8.1) con le procedure quantitative e statistiche che permettono di interpretare i risultati in modo corretto.

Il processo di validazione inizia già durante le prime fasi di sviluppo del test (paragrafo 2.7), nel momento in cui si decidono scopi e usi del test e si espongono le inferenze che si vorrebbero poter dedurre. Il secondo passo consiste nella raccolta di prove (*evidence*) che supportino la validità. Per

constatare la validità di un test è necessario effettuare la raccolta di tutta una serie di dati che consentono di interpretare correttamente i punteggi ottenuti dalla somministrazione del test.

Le prove raccolte possono essere frutto di analisi di contenuto in termini di *item* e di TLU, analisi di costrutto e analisi di criterio.

L'analisi di contenuto può essere effettuata coinvolgendo nella valutazione più correttori. In tal caso è possibile misurare il grado di accordo tra essi sulla valutazione di un campione di *item*, scelto come rappresentante adeguato dell'abilità da misurare, mediante la formula che segue:

Formula n. 3.65:
$$VC = \frac{N_{CA} - N_C/2}{N_C/2}$$

in cui la **validità di contenuto** (VC) è data dalla differenza tra il numero di correttori che si trovano in accordo (N_{CA}) meno la metà del numero totale di correttori (N_C), diviso la metà dei correttori coinvolti.

Una validità positiva va da 0,50 a 0,99, ossia si ottiene nel momento in cui più della metà dei valutatori giudica l'*item* essenziale e adeguato allo scopo.

L'analisi di costrutto può orientarsi verso la raccolta di prove sull'omogeneità del test, ossia sulla coerenza interna tra gli *item* che compongono il test, e sulla misurazione di un costrutto mediante il calcolo del coefficiente *alpha* di Cronbach o le formule n. 20 e n. 21 elaborate da Kuder e Richardson.

L'analisi di criterio può essere effettuata mediante il calcolo del coefficiente di correlazione, qualora si tratti di validità concorrente, e tramite la regressione, qualora si tratti di validità predittiva.

Proprio perché la validità di un test può essere valutata mediante

strumenti statistici, primo fra tutti la correlazione tra due serie di punteggi, il coefficiente di correlazione viene indicato spesso anche come “coefficiente di validità” ed esprime il grado di validità del test.

3.3.5 Interpretazione dei punteggi

I punteggi ottenuti da un test non posseggono un valore monosemico e necessitano di una reinterpretazione e organizzazione adeguata allo scopo per cui il test è progettato.

Spesso da un test si ottengono punteggi grezzi, i quali non si prestano a una facile interpretazione da parte degli utenti, soprattutto per coloro che non hanno familiarità con il contenuto del test, il numero di *item*, le procedure di correzione, le scale di valutazione, ecc.

Inoltre il punteggio finale, di frequente, è un punteggio composito, ossia la somma di svariati punteggi ricavati dalle molteplici parti del test.

I punteggi grezzi non possono essere interpretati direttamente, occorre innanzitutto verificare la scala di misurazione, poiché se si tratta di un punteggio composito, le scale di misurazione di ogni parte del test possono essere differenti e i punteggi grezzi non possono essere confrontati. Si necessita di un'unica scala di misurazione o di conversioni affinché si possano disporre i punteggi su una scala che abbia lo stesso punto di partenza e le stesse unità di misura.

Per tali ragioni è preferibile trovare dei metodi per riportare i dati in modo chiaro e organizzato e poter fornire un mezzo semplice e funzionale a coloro che vogliono utilizzarli e interpretarli (candidati, insegnanti, parenti, agenzie di impiego, sviluppatori del test, ecc.).

Per i test criteriali è possibile convertire i punteggi grezzi in valutazioni che possono essere interpretate in termini di grado di padronanza in una determinata area di competenza o livello di abilità raggiunto, in base ai criteri e domini prefissati.

Il grado o livello di padronanza raggiunto può essere interpretato mediante descrittori (*Can Do statements*) che indicano cosa il candidato è in grado di fare con la lingua in un dato livello oppure attraverso punteggi percentuali.

I punteggi di tipo criteriale possono essere calcolati e interpretati come **percentuale di risposte corrette** ($X_{\%}$) mediante la formula seguente:

$$\text{Formula n. 3.66:} \quad X_{\%} = \left(\frac{X}{K} \right) 100$$

in cui X indica il punteggio grezzo e K indica il numero degli *item* del test.

Nel caso di domande a credito parziale si applica la seguente formula:

$$\text{Formula n. 3.67:} \quad X_{\%} = \left(\frac{X}{X_{\max}} \right) 100$$

in cui X indica il punteggio grezzo e X_{\max} indica il punteggio massimo ottenibile nel test.

Queste formule permettono di convertire i punteggi grezzi in punteggi percentuali (dunque su una stessa scala ordinale), indicando lo 0 come punto di partenza.

Per i test di tipo normativo i punteggi grezzi vengono convertiti in valutazioni che forniscono informazioni sul livello del candidato rispetto al gruppo a cui appartiene o rispetto al *norm group* (paragrafo 2.9).

I punteggi dei test normativi possono essere riportati mediante graduatorie, ossia mediante liste di punteggi ordinate dal più alto al più basso. La

graduatoria costituisce una scala ordinale che, in quanto tale, non fornisce informazioni sulla distanza tra una posizione e l'altra. Inoltre l'interpretazione delle posizioni in graduatoria dipende dal numero dei candidati coinvolti.

Il metodo più diffuso è così il **rango percentile**, il quale indica la percentuale di candidati appartenenti a un gruppo i cui punteggi sono uguali o al di sotto di un dato punteggio. Ad esempio il punteggio di un candidato che è al ventesimo percentile indica che il 20% degli studenti dello stesso gruppo ha un punteggio inferiore o uguale a quello del candidato del ventesimo percentile. Il rango percentile si calcola mediante la formula che segue:

$$\text{Formula n. 3.68:} \quad R \%ile = \left(\frac{cumf}{Totf} \right) 100$$

in cui *cumf* è la frequenza cumulativa e *Totf* è la frequenza totale.

La frequenza cumulativa si calcola, partendo dal punteggio più basso, aggiungendo la sua frequenza a quella del punteggio successivo, e così per ogni punteggio a salire. La frequenza cumulativa del punteggio più alto equivale sempre alla frequenza totale.

Ad esempio, considerando i punteggi presentati nella tabella n. 3.1 - B, la *cumf* per ogni punteggio sarà:

Punteggio <i>x</i>	Frequenza <i>f</i>	<i>cumf</i>
40	1	15
37	1	14
36	2	13
35	1	11
34	1	10
33	1	9
32	3	8
31	2	5
29	1	3
24	1	2
17	1	1

Tabella n. 3.7 - Esempio di calcolo della frequenza cumulativa

Dopo il calcolo della *cumf*, è possibile calcolare il rango percentile per ciascun punteggio utilizzando la formula n. 3.68 e arrotondando il punteggio al numero intero più prossimo, per eccesso o per difetto in base ai casi:

Punteggio <i>x</i>	Frequenza <i>f</i>	<i>cumf</i>	Rango percentile
40	1	15	100
37	1	14	93
36	2	13	87
35	1	11	73
34	1	10	67
33	1	9	60
32	3	8	53
31	2	5	33
29	1	3	20
24	1	2	13
17	1	1	7

Tabella n. 3.8 - Esempio di calcolo del rango percentile (Bachman, 2004:307)

Il rango percentile non va confuso con il calcolo della percentuale di risposte corrette ($X_{\%}$), poiché il primo indica la posizione del candidato rispetto al gruppo, mentre il secondo indica il grado di padronanza raggiunto dal candidato in una determinata area di competenza della lingua rispetto al criterio.

Rispetto alla graduatoria, l'interpretazione del rango percentile non dipende dal numero dei candidati coinvolti nella misurazione. Inoltre presenta il vantaggio della facilità di calcolo e di interpretazione. Nonostante ciò, come per la graduatoria, anch'esso si presenta in forma di scala ordinale, dunque non permette la distinzione tra un percentile e l'altro in termini di differente livello di abilità.

Il metodo che supera il limite della scala ordinale è il punteggio standard (calcolabile mediante la formula n. 3.12), il quale fornisce un'unica scala a intervalli equidistanti e basati su una unità di misura.

3.4 Riflessioni conclusive

Lo scopo di questo capitolo consiste nel presentare e chiarire alcune procedure statistiche che permettono di dare prova empirica dei risultati conseguiti nei test.

La statistica, nel campo delle scienze umane, ampiamente applicata ai test psicologici, è ancora poco impiegata nell'analisi dei test linguistici. Lo scopo di questa ricerca è quello di facilitare, attraverso una chiara descrizione di alcune procedure statistiche, il lavoro di analisi e interpretazione dei punteggi e di agevolare il lavoro degli *stakeholder* della formazione e della valutazione nell'applicazione di tali procedure, con il fine ultimo di incoraggiare e diffondere l'utilizzo di metodi statistici nel campo del *testing* italiano, poiché si ritiene che tale tipologia di indagine possa offrire prove quantitative pratiche per il miglioramento dei test e dei sistemi valutativi.

In questo capitolo si è scelto di illustrare alcune formule che appartengono alla statistica descrittiva e che conducono alla formulazione di inferenze sulle competenze di specifici soggetti sottoposti a esame, piuttosto che formule di tipo inferenziale, attraverso le quali, mediante un'indagine su un campione, è possibile dedurre le competenze di un'intera popolazione.

Nel presente lavoro la scelta è caduta sulle procedure di tipo descrittivo per due ragioni: la prima è data dal fatto che la statistica descrittiva pone le basi per quella inferenziale e, dunque, è di primaria importanza; la seconda motivazione consiste nel poter offrire un esempio applicativo di analisi statistica di tipo descrittivo su un caso concreto (presentato nel capitolo successivo) che possa fungere da esempio per applicazioni pratiche in contesti didattici di piccole e/o medie dimensioni, ossia laddove le figure di docente, sviluppatore del test e

valutatore in genere coincidono e presentano ancora poca dimestichezza con tali tecniche. L'elaborazione statistica dei dati di un test, dopo le prime applicazioni delle formule, diventa semplice; e pertanto si ritiene che un'attiva costante di *testing* in questo senso potrebbe portare a interessanti sviluppi nel campo e potrebbe rientrare nelle "buone pratiche" della valutazione linguistica.

Un caso di studio: PFLISP Sicilia

4.1 Premesse introduttive

Il presente capitolo illustra l'applicazione di alcune procedure statistiche descritte nel precedente capitolo su un caso concreto di realtà valutativa. Lo scopo è quello di fornire uno strumento pratico ed esemplificativo per eseguire analisi quantitative sui dati di un test e per interpretare correttamente i risultati ottenuti.

Gli strumenti statistici forniscono infatti prove per avvalorare empiricamente previsioni e ipotesi sulle competenze e abilità linguistiche dei candidati, previo esame dell'affidabilità e validità dei test progettati, sviluppati e utilizzati per raggiungere un determinato obiettivo. I dati ottenuti forniscono un efficace aiuto per migliorare tali test e dunque poter prendere decisioni sui candidati in base allo scopo per cui son stati sottoposti a esame.

Le procedure qui analizzate vengono applicate a un caso concreto per dar prova della loro efficacia e fornire uno strumento didattico utile a coloro i quali si interessano di *testing* linguistico o sono coinvolti in qualche modo in processi di valutazione linguistica.

4.2 Il Progetto PFLISP Sicilia

Il progetto di applicazione delle formule statistiche e di analisi dei risultati su un caso concreto nasce dall'esigenza di avere un quadro più chiaro della qualità di test linguistici utilizzati per il progetto PFLISP Sicilia e dell'andamento delle *performance* dei candidati sottoposti a tali prove.

Il Progetto PFLISP Sicilia parte dalla proposta avanzata dall'Ufficio Scolastico Regionale della Sicilia al CLA di Messina, al CLA di Palermo, al CLIK di Enna e al CLMA di Catania di redigere le prove concorsuali e valutare le *performance* dei candidati sottoposti all'esame conclusivo previsto dal "*Piano di Formazione per lo Sviluppo delle Competenze linguistico-comunicative e metodologico-didattiche in Lingua Inglese degli insegnanti di Scuola Primaria*" della Regione Sicilia.

Il "*Piano Nazionale di Formazione per lo Sviluppo delle Competenze linguistico-comunicative e metodologico-didattiche in Lingua Inglese degli insegnanti di Scuola Primaria*" (MIUR prot. n. 1446 del 19 luglio 2005¹) nasce in seguito alla Legge Finanziaria 2005² e alla Legge Finanziaria 2007³, le quali recitano rispettivamente:

"L'insegnamento della lingua straniera nella scuola primaria è impartito dai docenti della classe in possesso dei requisiti richiesti o da altro docente facente parte dell'organico di istituto sempre in possesso dei requisiti richiesti. [...] Al fine di realizzare quanto previsto dal presente comma, [...],

¹ Il testo del protocollo n. 1446 del MIUR è reperibile online all'indirizzo: http://archivio.pubblica.istruzione.it/normativa/2005/prot1446_05.shtml (ultima consultazione: febbraio 2011).

² Il testo della Legge n. 311 del 30 dicembre 2004 (Finanziaria 2005) è consultabile online all'indirizzo: <http://www.parlamento.it/elenchileggi/43858/75870/elencoleggi.htm> (ultima consultazione: febbraio 2011).

³ Il testo della Legge n. 296 del 27 dicembre 2006 (Finanziaria 2007) è consultabile online all'indirizzo: <http://www.parlamento.it/elenchileggi/43858/44080/elencoleggi.htm> (ultima consultazione: febbraio 2011).

sono attivati corsi di formazione, nell'ambito delle annuali iniziative di formazione in servizio del personale docente, la cui partecipazione è obbligatoria per tutti i docenti privi dei requisiti previsti per l'insegnamento della lingua straniera [...]. (311/04, art. 128)

“ai fini della compiuta attuazione di quanto previsto dall'articolo 1, comma 128, della legge 30 dicembre 2004, n. 311, l'adozione di un piano biennale di formazione per i docenti della scuola primaria, da realizzare negli anni scolastici 2007/2008 e 2008/2009, finalizzato al conseguimento delle competenze necessarie per l'insegnamento della lingua inglese. A tale fine, per un rapido conseguimento dell'obiettivo, sono attivati corsi di formazione anche a distanza, integrati da momenti intensivi in presenza”.

(296/06, art. 605 lettera e)

Le Leggi di cui sopra nascono in seguito al Decreto Legislativo n. 59/04⁴, attuativo della Legge n. 53/03⁵, sulla definizione delle norme generali relative alla scuola dell'infanzia e al primo ciclo dell'istruzione, che prevede l'insegnamento della lingua inglese come parte integrante del percorso di formazione dell'alunno già dal primo anno della scuola primaria.

Il Piano Nazionale deriva dall'esigenza, in seguito alla decisione di inserire *“l'insegnamento della lingua inglese [come] parte integrante del percorso complessivo di formazione dell'alunno sin dal primo anno di scolarità [il quale deve essere] impartito da insegnanti in base al principio dell'unitarietà del sapere”* (MIUR prot. n. 1446), di attivare interventi di formazione linguistico-comunicativa in favore di docenti di scuola primaria privi dei requisiti necessari per l'insegnamento della lingua inglese, attraverso la pianificazione, nonché l'attuazione, di un'azione sistematica e pluriennale di formazione in ambito linguistico, didattico e metodologico. Tale progetto di formazione degli insegnanti

⁴ Il testo del Decreto Legislativo n. 59 del 19 febbraio 2004 è reperibile online all'indirizzo: <http://www.parlamento.it/parlam/leggi/deleghe/04059dl.htm> (ultima consultazione: febbraio 2011).

⁵ Il testo della Legge n. 53 del 28 marzo 2003 è reperibile online all'indirizzo: <http://www.parlamento.it/elenchileggi/43858/75871/elencoleggi.htm> (ultima consultazione: febbraio 2011).

di lingua va interpretato alla luce delle innovazioni in corso nel sistema scolastico.

Il Piano, coordinato dall'Agenzia Nazionale per lo Sviluppo dell'Autonomia Scolastica (ex INDIRE – Istituto Nazionale di Documentazione per l'Innovazione e la Ricerca Educativa), è regolamentato dal Documento a cura del Comitato Tecnico Scientifico costituito presso l'INDIRE⁶, il quale fissa gli obiettivi formativi da raggiungere, l'organizzazione e le caratteristiche dei corsi, il monitoraggio in itinere, la valutazione delle competenze e il profilo professionale del docente di lingua inglese.

Tale progetto assume grande rilevanza se si considera, come scrive Gotti, “che l'ampliamento dell'offerta didattica di lingue straniere determinato dall'applicazione della riforma si troverà spesso a dover fare i conti con una situazione caratterizzata dalla mancanza di adeguate competenze linguistico-comunicative, senza le quali qualsiasi tentativo di insegnare una lingua straniera risulta problematico e destinato al fallimento” (Gotti, *Proposta per la formazione linguistica dei docenti*⁷).

Poiché gli insegnanti di lingua inglese presenti nella scuola primaria si presentano con livelli di competenza diversi, il programma si prefigge l'armonizzazione delle componenti linguistico-comunicative dei docenti e la loro progressiva formazione linguistica.

Il piano qui delineato ha l'obiettivo principale di condurre ogni docente di lingua al possesso delle competenze linguistico-comunicative di base. Si tratta di far raggiungere il livello considerato di base, identificato nel B1 del QCER per la scuola d'infanzia e primaria, “che garantisce un'adeguata padronanza d'uso della

⁶ Tale documento è reperibile tra gli allegati del prot. n. 1446 sotto il titolo di “Formazione linguistica 2005”.

⁷ Tale documento è reperibile tra gli allegati del prot. n. 1446 sotto il titolo di “Formazione linguistica 2004”.

lingua straniera oggetto di insegnamento, nonché la capacità di riflettere sul sistema linguistico di riferimento e di descriverne e analizzarne il funzionamento nella realizzazione delle sue principali funzioni comunicative” (Gotti, *Proposta per la formazione linguistica dei docenti*). Tale livello di competenza linguistica infatti, affinché si possa garantire un apprendimento precoce delle lingue straniere nell’alunno in maniera efficace, dovrebbe diventare un requisito indispensabile per l’accesso all’insegnamento.

I percorsi formativi previsti dal Piano Nazionale riguardano processi diversificati in base all’utenza e alle esigenze territoriali di ciascuna regione. Essi possono rispondere a esigenze di alfabetizzazione linguistica, di recupero delle conoscenze pregresse e di potenziamento delle conoscenze esistenti, aventi tutti l’obiettivo ultimo di portare i destinatari al raggiungimento del predetto livello B1, requisito minimo richiesto dal profilo professionale⁸.

I destinatari di tale processo di formazione sono individuati tra i docenti generalisti con contratto a tempo indeterminato che non insegnano la lingua inglese e che volontariamente dichiarano la propria disponibilità a seguire tali corsi di formazione. Tra i vari candidati vengono privilegiati coloro che dichiarano una competenza minima di livello A2 e che insegnano lingue straniere diverse dall’inglese. Vengono suggeriti⁹, dunque, test informatizzati di autovalutazione per individuare il proprio livello di competenza linguistica secondo i parametri del QCER (il più comune è il *Quick Placement Test*¹⁰ dell’Università di Oxford e

⁸ Il “Profilo professionale del docente di lingua inglese nelle scuole dell’infanzia e primaria” è allegato al prot. n. 1446.

⁹ Per maggiori dettagli si rimanda alla consultazione dei documenti: “Linee di orientamento per la formazione n servizio dei docenti di lingua inglese delle scuole dell’infanzia e primaria” e “Proposta per la formazione linguistica dei docenti”, allegati al prot. n. 1446.

¹⁰ Il *Quick Placement Test* è disponibile online all’indirizzo:<http://www.britishcouncil.org/lithuania-education-uk-exams-in-lithuania-english-exams-placement.htm> (ultima consultazione: febbraio 2011) oppure <http://www.oxfordenglishtesting.com/DefaultMR.aspx?id=3034&menuId=1#> (ultima

Cambridge).

Gli interventi formativi previsti dal Piano Nazionale riguardano la messa a punto di corsi in presenza, modulari e diversificati per livello, di durata variabile in base al livello di competenza accertato inizialmente. Infatti il Piano, nonché i relativi allegati, prevedono la somministrazione di un test d'ingresso sulla base dei descrittori del QCER, di competenza dell'Agenzia Nazionale per lo Sviluppo dell'Autonomia Scolastica, per l'accertamento delle competenze in ingresso dei docenti.

La formazione linguistica deve essere integrata con corsi di formazione per lo sviluppo di competenze metodologico-didattiche erogati, in modalità *e-learning*, anche in questo caso dall'Agenzia Nazionale per lo Sviluppo dell'Autonomia Scolastica.

I corsi di formazione linguistica prevedono una valutazione finale e il rilascio di un attestato che certifichi il percorso svolto e il livello di competenza linguistico-comunicativa raggiunto secondo il QCER. Le scale comuni di riferimento stabilite dal QCER forniscono infatti una valida base per descrivere le competenze linguistiche degli individui in maniera obiettiva, trasparente e standardizzata. L'acquisizione di una certificazione riconosciuta a livello internazionale è da ritenersi dunque un'opportunità di omogeneità di preparazione su tutto il territorio nazionale e una garanzia sullo standard delle competenze raggiunte. Attraverso tale certificazione il docente può confrontarsi con parametri stabiliti a livello nazionale e internazionale, sviluppando un proprio percorso professionale validato, e può spendere tale certificazione anche in situazioni extra scolastiche.

consultazione: febbraio 2011).

La valutazione finale e il rilascio della certificazione vengono affidati in via prioritaria alle Università e ai Centri Linguistici di Ateneo in accordo con l'attuale sistema istituzionale, il quale favorisce, per la formazione, le strutture universitarie.

Ai Centri Linguistici di Ateneo è affidato anche il compito di organizzare i corsi di lingua in presenza.

La scelta dei formatori per lo svolgimento dei corsi di formazione è guidata da criteri stabiliti dal Piano Nazionale, i quali prevedono il possesso di una competenza linguistico-comunicativa di livello C1 del già richiamato QCER e del titolo di specializzazione in ELT (*English Language Teaching*), un'attività pluriennale di insegnamento della lingua inglese, un'esperienza di formazione linguistica ad adulti e una formazione acquisita in progetti nazionali e/o internazionali.

Ogni Ufficio Scolastico Regionale è tenuto a predisporre di un proprio Piano di formazione basato sul Piano Nazionale e sui documenti sopraindicati.

Il Progetto Regionale per la Sicilia PFLISP¹¹ si pone come finalità primarie l'aumento del numero dei docenti generalisti con contratto a tempo indeterminato che insegnano anche la lingua inglese, il raggiungimento per questi docenti almeno del livello B1 del QCER e la rilevazione in uscita delle competenze linguistico-comunicative acquisite.

I formatori di lingua, in seguito alla presentazione di un'autocandidatura, sono selezionati mediante apposito bando regionale e giudicati da un'opportuna commissione istituita dall'Ufficio Scolastico Regionale. Tali formatori, scelti tra

¹¹ Il testo del Progetto Regionale per la Sicilia PFLISP è disponibile online all'indirizzo:
http://62.77.63.181/ISN_Istruionesicilia_it/pagina.aspx?pageid=a48b0bcd-6404-486c-89e6-d85d50d38426 (ultima consultazione: febbraio 2011).

docenti in servizio presso le istituzioni scolastiche, formatori che hanno collaborato al progetto Lingue 2000, esperti suggeriti da Università e Centri Linguistici, ed esperti in servizio presso le scuole di lingua accreditate, sono tenuti a stabilire e concordare insieme metodi e strumenti da utilizzare per la valutazione delle competenze iniziali dei corsisti e per la costituzione delle classi.

Il Progetto PFLISP Sicilia prevede la costituzione di corsi per principianti (della durata di 380 ore distribuite nell'arco di due anni) e corsi intermedi (della durata di 280 ore distribuite nell'arco di due anni e 160 in un anno, in base al livello di partenza di ciascun corsista).

L'Ufficio Scolastico Regionale per la Sicilia attribuisce ai quattro CLA prima menzionati il compito di elaborare le prove per la certificazione finale e di effettuare la relativa valutazione. Per l'elaborazione delle prove i quattro CLA delle Università di Catania, Enna, Messina e Palermo delineano un piano (proposto dagli esperti linguistici e approvato dal rappresentante dello USR Sicilia nonché referente scientifico del progetto) per l'elaborazione, lo svolgimento e la valutazione delle prove al fine del rilascio della certificazione finale delle competenze linguistico-comunicative del PFLISP Sicilia.

4.2.1 Fase I: elaborazione delle prove

Per l'elaborazione delle prove lo staff è stato scelto tra esperti linguistici che collaborano con i quattro CLA siciliani e/o dai formatori dei corsi.

Gli esperti hanno elaborato prove di verifica identiche per tutti i candidati della Regione Sicilia attenendosi al principio dell'unicità dei contenuti.

Le prove rispettano il *format* dell'esame Cambridge per il livello B1 del

QCER, ossia il PET (*Preliminary English Test*) secondo la scala del Cambridge ESOL (paragrafo 1.8), che riguarda l'area del *General English*.

L'esame PET consiste in tre prove: la prima (90 minuti) è divisa in esercizi di lettura (5 parti) e scrittura (3 parti); la seconda (30 minuti) è una prova di ascolto (4 parti); la terza (10-12 minuti) è un colloquio orale (4 parti).

Le prove sono così composte:

Abilità	Comprensione scritta
Tipo di esercizio	<ol style="list-style-type: none"> 1. 5 <i>item</i> a scelta multipla a tre opzioni su cinque brevi testi; 2. 5 <i>item</i> di abbinamento tra brevi testi autentici semplificati e descrizioni di persone; 3. 10 <i>item</i> vero/falso su un lungo testo autentico semplificato; 4. 5 <i>item</i> a scelta multipla a quattro opzioni relative a un lungo testo autentico semplificato; 5. 10 <i>item</i> a scelta multipla a quattro opzioni in un <i>cloze test</i>.
Modalità di somministrazione	Supporto cartaceo o informatico
Tempo	90' (compreso la prova di scrittura)
Punteggio	Ciascun <i>item</i> vale un punto, ma il totale influisce sull'intero esame per il 25%.

Tabella n. 4.1- Composizione della prova di lettura PET

Abilità	Produzione scritta
Tipo di esercizio	<ol style="list-style-type: none"> 1. 5 <i>item</i> da trasformare mantenendo il significato originale; 2. redazione di un breve messaggio (cartolina, email, nota) di 35-45 parole; 3. redazione di un lungo testo (storia o lettera informale) di circa 100 parole.
Modalità di somministrazione	Supporto cartaceo o informatico
Tempo	90' (compreso la prova di lettura)
Punteggio	25 di cui 5 punti per la prima parte, 5 punti per la seconda parte, 15 punti per la terza parte (25% dell'intero esame)

Tabella n. 4.2- Composizione della prova di scrittura PET

Abilità		Comprensione orale
Tipo di esercizio		<ol style="list-style-type: none"> 1. 7 <i>item</i> a scelta multipla a tre opzioni visive su brevi monologhi o dialoghi; 2. 6 <i>item</i> a scelta multipla a tre opzioni su un lungo monologo o un'intervista; 3. 6 <i>item</i> di riempimento relativi a un lungo monologo; 4. 6 <i>item</i> vero/falso su un lungo dialogo informale.
Modalità di somministrazione		Supporto cartaceo o informatico
Tempo		30'
Punteggio		25 (25% dell'intero esame)

Tabella n. 4.3- Composizione della prova di ascolto PET

Abilità		Produzione orale
Tipo di esercizio		<ol style="list-style-type: none"> 1. interazione tra ciascun candidato e l'esaminatore (2-3 minuti); 2. interazione tra i candidati partendo da uno stimolo visivo 2-3 minuti); 3. monologo su una foto (1 minuto per ciascun candidato); 4. conversazione tra i candidati sul monologo della parte 3 (3 minuti).
Modalità di somministrazione		Conversazione orale
Tempo		10'- 12'
Punteggio		25 (25% dell'intero esame)

Tabella n. 4.4- Composizione della prova di parlato PET

La prova di lettura attesta la capacità di leggere e comprendere testi autentici, tratti dal mondo reale, da quotidiani e riviste, brochure e manifesti.

La prova di scrittura attesta la capacità di comporre testi semplici e coerenti, di controllare le strutture grammaticali, il vocabolario, l'ortografia e la punteggiatura.

La prova di ascolto valuta la capacità di comprendere materiale registrato tratto da situazioni reali, in cui sono presenti parlanti madrelingua con varianti linguistiche standard diverse.

La prova orale valuta la capacità di interagire in una conversazione, saper rispondere a domande, porre quesiti e parlare liberamente dei propri pensieri e

interessi.

Il punteggio finale è ottenuto dalla somma dei punteggi conseguiti nelle tre prove. I candidati che superano l'esame ricevono un certificato con l'indicazione del voto: se il punteggio è tra 85 e 100 il giudizio è espresso con la valutazione "Pass with Merit" (promosso con merito), se il punteggio è tra 70 e 84 con "Pass" (promosso). I candidati che ottengono un punteggio tra 45 e 69 ottengono un certificato A2, poiché dimostrano una conoscenza linguistica di livello inferiore; diversamente coloro che ottengono un punteggio tra 0 e 44 non superano l'esame (il giudizio è espresso con la valutazione "Fail").

L'esame per il progetto PFLISP Sicilia consiste in quattro prove: la prima (30 minuti) è una prova di lettura (3 parti), la seconda (30 minuti) è una prova di ascolto (2 parti), la terza (45 minuti) è una prova di scrittura (2 parti), la quarta (10 minuti) è un colloquio orale (1 parte).

Le prove sono così composte:

Abilità	Comprensione scritta
Tipo di esercizio	<ol style="list-style-type: none"> 1. 5 <i>item</i> a scelta multipla a tre opzioni relative a un brano; 2. 5 <i>item</i> a scelta multipla a tre opzioni relative a un brano; 3. 5 <i>item</i> vero/falso su un testo; 4. 10 <i>item</i> a scelta multipla a tre opzioni in un <i>cloze test</i>.
Modalità di somministrazione	Supporto informatico
Tempo	30'
Punteggio	25

Tabella n. 4.5- Composizione della prova di lettura PFLISP

Abilità	Comprensione orale
Tipo di esercizio	<ol style="list-style-type: none"> 1. 5 <i>item</i> a scelta multipla a tre opzioni su cinque brevi monologhi o dialoghi; 2. 5 <i>item</i> a scelta multipla a tre opzioni su un lungo monologo o dialogo.
Modalità di somministrazione	Supporto informatico
Tempo	30'
Punteggio	20

Tabella n. 4.6- Composizione della prova di ascolto PFLISP

Abilità	Produzione scritta
Tipo di esercizio	1. 10 <i>item</i> da trasformare mantenendo il significato originale; 2. redazione guidata di un elaborato (lettera, breve racconto o resoconto, etc.) di circa 100-120 parole.
Modalità di somministrazione	Supporto cartaceo
Tempo	45'
Punteggio	25 di cui 10 punti per la prima parte, 15 punti per la seconda parte (massimo 5 punti per la correttezza formale e massimo 10 punti per contenuto e coerenza)

Tabella n. 4.7- Composizione della prova di scrittura PFLISP

Abilità	Produzione orale
Tipo di esercizio	1. conversazione con l'esaminatore su argomenti compresi nei <i>domain Personal/Public/Educational</i> del QCER.
Modalità di somministrazione	Colloquio
Tempo	10'
Punteggio	30 (massimo 10 punti per la correttezza formale; massimo 10 punti per la competenza comunicativa; massimo 10 punti per il lessico).

Tabella n. 4.8- Composizione della prova di parlato PFLISP

Per assicurare l'omogeneità in merito alla formazione e alla valutazione delle competenze linguistiche su tutto il territorio nazionale, i Centri Linguistici di Ateneo si attengono a strumenti e livelli indicati dal QCER. Pertanto le prove sono finalizzate all'attestazione del possesso delle competenze linguistico-comunicative così come sono specificate, per il livello B1, nella griglia di autovalutazione del QCER (tabella n. 1 in appendice).

Inoltre gli argomenti delle quattro prove vertono sui tre domini o sfere d'interesse previsti nel QCER per il livello B1: Privato, Pubblico e Educativo (tab. n. 2 in appendice); in particolare per le prove di lettura e ascolto almeno il 50% degli argomenti deve essere tratto dal dominio educativo.

Il punteggio finale è ottenuto dalla somma dei punteggi conseguiti nelle quattro prove. L'attribuzione del punteggio però non è equilibrata, poiché la

soglia di superamento della prova è posta a 65/100¹², ma l'esame può essere considerato superato solo se nella prova di produzione orale si raggiunge un punteggio non inferiore a 20/30.

La rilevanza attribuita alle prove nel PFLISP e nel PET è piuttosto simile:

	Letture	Ascolto	Scrittura	Colloquio
PFLISP	25%	20%	25%	30%
PET	25%	25%	25%	25%

Tabella n. 4.9 - Rilevanza prove livello B1

così come per il tempo concesso per lo svolgimento delle singole prove:

	Letture	Ascolto	Scrittura	Colloquio
PFLISP	30'	30'	45'	10'
PET	90' (comprensivo del tempo concesso per la prova di scrittura)	30'	90' (comprensivo del tempo concesso per la prova di lettura)	10' - 12'

Tabella n. 4.10 – Tempo concesso per le prove di livello B1

La diversità relativa al tempo, riscontrabile nelle prove di lettura e scrittura, è una conseguenza del diverso numero di *item* proposti e della quantità di lavoro richiesta al candidato per ogni prova.

Le difformità rispetto al *format* PET riguardanti la tipologia di *item*, il numero di *item* e il peso delle singole prove scaturiscono dall'accordo tra i quattro CLA, i formatori linguistici e l'Ufficio Scolastico Regionale per la Sicilia. La ragione di tali difformità è attribuibile allo scopo della certificazione e allo specifico target a cui sono destinate le prove.

Esse si basano soprattutto sul *format* stabilito dalle linee guida presentate nel documento ministeriale: “*Quale profilo e quali competenze per l'inglese del docente di scuola primaria - Un'esperienza di ricerca*”¹³, nel quale vengono

¹² La soglia è stata stabilita a seguito di incontri e accordi raggiunti con l'Ufficio Scolastico Regionale per la Sicilia.

¹³ Il documento “*Quale profilo e quali competenze per l'inglese del docente di scuola primaria - Un'esperienza di ricerca*” è disponibile online all'indirizzo: <http://puntoeduri.indire.it/lingue> (ultima

presentate alcune prove esemplificative provenienti da vari CLA e dal CLA di Trieste in particolare. Attraverso l'esame di tali prove si può constatare un alto livello di omogeneità e standardizzazione dei test a livello nazionale.

Al superamento dell'esame i CLA rilasciano una certificazione, la cui validità rientra nel Quadro della convenzione tra il Ministero dell'Istruzione, dell'Università e della Ricerca attraverso INDIRE e l'AICLU (Associazione Italiana Centri Linguistici Universitari)¹⁴.

4.2.2 Fase II: attuazione delle prove

L'esame PFLISP prevede la somministrazione delle prove mediante supporti diversi: le due abilità ricettive, *reading* e *listening*, attraverso supporto informatico e le due abilità produttive, *writing* e *speaking*, rispettivamente attraverso supporto cartaceo e colloquio individuale.

In questo lavoro si descriverà lo svolgimento e la valutazione delle prove che hanno avuto luogo presso il CLA dell'Università di Palermo e si analizzeranno i risultati relativi ai candidati della suddetta provincia.

Per la verifica delle abilità di comprensione scritta e comprensione orale l'esame è stato reso possibile attraverso la piattaforma Moodle¹⁵ "Anacleto"¹⁶ del Corso di Laurea in "Lingue Moderne per il Web" della Facoltà di Lettere e

consultazione: febbraio 2011).

¹⁴ Il documento Protocollo di intesa tra INDIRE e AICLU è disponibile online all'indirizzo: http://archivio.pubblica.istruzione.it/normativa/2006/allegati/prot233_06.pdf (ultima consultazione: febbraio 2011).

¹⁵ Moodle è una piattaforma per la formazione a distanza, ossia un pacchetto software per erogare e gestire corsi di formazione *on-line*.

¹⁶ La piattaforma "Anacleto" è disponibile online all'indirizzo: <http://147.163.11.3/anacleto2008> (ultima consultazione: febbraio 2011).

Filosofia dell'Università degli Studi di Palermo.

Ai candidati è stata offerta la possibilità di accedere a un *sample test* per familiarizzare con il *format*. Inoltre è stata offerta loro la possibilità di acquisire dimestichezza con lo strumento informatico attraverso momenti di laboratorio liberi e facoltativi (monitorati dal personale CLA) per l'esercizio delle prove ricettive.

Gli esami hanno avuto luogo al termine dei corsi di formazione linguistica nel periodo compreso tra marzo e giugno 2009. I candidati, suddivisi in gruppi, hanno svolto le prove presso le aule e i laboratori del CLA di Palermo.

Ogni candidato è stato sottoposto all'esame in due momenti diversi: nella prima giornata hanno avuto luogo le prove di lettura, ascolto e scrittura (nell'ordine in cui sono qui elencate) e, a distanza di una settimana circa, in una seconda giornata, si è svolta la prova orale.

Per le prove di comprensione orale i candidati hanno potuto ascoltare i brani tre volte e gestire la registrazione liberamente (fermare la registrazione, tornare indietro, avanzare, ecc.).

4.2.3 Fase III: valutazione delle prove

Le prove sono state corrette e valutate da una commissione costituita dagli stessi esperti linguistici che hanno preparato i test.

Per le prove di comprensione scritta e orale, somministrate mediante supporto informatico, la correzione è stata automatizzata e computerizzata. Si tratta infatti di *item* di tipo oggettivo, per cui a ogni risposta corretta è assegnato 1 punto, e ogni risposta errata o omessa equivale a 0. Non vi sono risposte con

punteggio parziale e penalizzazione delle risposte errate come per la certificazione PET.

Nonostante lo strumento informatico fosse in grado di fornire un *feedback* immediato in merito all'esito delle prove ricettive, si è scelto di non comunicare la valutazione immediatamente, poiché i candidati, dopo le prove suddette, sono stati sottoposti alla prova di produzione scritta, e una valutazione negativa avrebbe potuto avere un'influenza negativa.

La valutazione delle tre prove è stata loro comunicata in un momento successivo alla prima giornata d'esame e comunque precedente alla seconda giornata d'esame.

4.3 Analisi dei dati del PFLISP Sicilia

Si è scelto di analizzare gli esiti delle prove di valutazione finale poiché necessitano di una dimostrazione quantitativa a supporto dell'efficacia qualitativa.

Diversamente i test di ingresso non sono stati sottoposti a esame statistico poiché, come si legge dal documento redatto da Gotti (*Proposta per la formazione linguistica dei docenti*), si tratta di prove elaborate da Enti linguistici e culturali accreditati (quale ad esempio il *Quick Placement Test* edito da Oxford University Press in collaborazione con UCLES), e si può sufficientemente essere sicuri che si tratti di test affidabili, validati secondo parametri scientifici, omogenei su tutto il territorio nazionale, oggettivi, corretti in maniera automatica e immediata, e infine sperimentati già da un vasto pubblico.

L'analisi riguarda i dati delle sole prove ricettive per le quali l'indagine statistica è svolta su dati oggettivi, in modo imparziale e automatico.

Gli esiti delle prove di valutazione finale elaborate e somministrate presso il CLA di Palermo sono analizzati in due blocchi separati, poiché i dati impiegati in questo studio sono stati raccolti durante due diverse sessioni d'esame (la prima nel marzo 2009 e la seconda nel giugno 2009). Le prove somministrate durante la seconda sessione sono l'esito di una rielaborazione basata sulla rilevazione statistica dei dati della sessione precedente.

È necessario precisare che i candidati sottoposti ai test sono stati suddivisi in 5 gruppi (A, B, C, D, E), ai quali sono state somministrate 5 prove diverse. Ogni candidato è distinto, per rispetto della *privacy*, mediante l'utilizzo dei numeri ordinali progressivi (1-154), in sostituzione dei propri dati anagrafici.

4.3.1 Analisi statistica degli *item*

L'analisi statistica degli *item* è stata condotta attraverso il calcolo degli indici di difficoltà e di discriminazione sulle cinque prove, sia di lettura che di ascolto. È stato adottato il metodo dei due gruppi con il punteggio rispettivamente più alto (*h*) e più basso (*l*), secondo il modello della teoria classica (paragrafo 3.3.2).

Indice di difficoltà

Per il calcolo dell'indice di difficoltà sono state impiegate le formule n. 3.24 $[IF = \frac{p_h + p_l}{2}]$ e/o n. 3.25 $[IF = \frac{R_h + R_l}{2n}]$ poiché le prove esaminate sono di tipo dicotomico, due punti per ogni risposta esatta per la prova di ascolto (max 20 punti), un punto per ogni risposta esatta per la prova di lettura (max 25 punti).

Prova di ascolto

Prova n. 1	IF valore accettabile 0,3 - 0,7	Prova n. 2	IF valore accettabile 0,3 - 0,7	Prova n.3	IF valore accettabile 0,3 - 0,7	Prova n. 4	IF valore accettabile 0,3 - 0,7
Parte prima							
Item 1	0,95	Item 1	0,84	Item 1	0,89	Item 1	0,89
Item 2	1	Item 2	0,84	Item 2	1	Item 2	0,89
Item 3	1	Item 3	1	Item 3	0,67	Item 3	1
Item 4	0,7	Item 4	0,84	Item 4	0,84	Item 4	1
Item 5	1	Item 5	0,78	Item 5	1	Item 5	1
Parte seconda							
Item 1	1	Item 1	0,95	Item 1	0,89	Item 1	0,89
Item 2	0,6	Item 2	0,89	Item 2	1	Item 2	0,84
Item 3	1	Item 3	0,84	Item 3	0,95	Item 3	1
Item 4	0,15	Item 4	0,67	Item 4	0,89	Item 4	0,95
Item 5	0,55	Item 5	0,45	Item 5	1	Item 5	0,62

Tabella n. 4.11 – Indice di difficoltà degli *item* delle prime quattro prove di ascolto

Per le prime quattro prove di ascolto, i valori IF che rientrano nella norma, contrassegnati nella tabella dal colore rosso, sono 7/40, quindi il 17,5%. L'80% degli *item* risulta troppo facile poiché i valori sono vicini al +1,00, il 2,5% troppo difficile poiché vicino allo 0.

La prima prova di ascolto, ad esempio, presenta 5 *item* con un indice di difficoltà pari a 1, dunque gli *item* sembrano essere troppo facili. Si riporta sotto l'esempio di uno degli *item* che presenta la situazione appena descritta:

Tipo di prova: ascolto		Gruppo: A		Totale candidati: 20		Data: 31 marzo 2009	
N. <i>item</i>	Risp.	Candidati gruppo A _i : 10	p	Candidati gruppo A _i : 10	p	IF	
Parte seconda							
1	a	-	0	-	0	0	
	b	10	1	10	1	1	
	c	-	0	-	0	0	
	O	-	0	-	0	0	

Tabella n. 4.12 - Prova di ascolto n. 1 Parte II *Item* n. 1

Dai dati della tabella si evince che sia i candidati migliori sia i meno competenti hanno risposto correttamente (per il quadro di tutte le risposte per ogni *item* di ciascuna prova si rimanda all'appendice).

La quinta prova, somministrata durante la seconda sessione, dunque elaborata in un momento successivo all'analisi statistica, presenta una situazione totalmente diversa, con risultati interessanti:

Prova n. 5	IF valore accettabile 0,3 - 0,7
Parte prima	
Item 1	0,2
Item 2	0,5
Item 3	1
Item 4	0,8
Item 5	0,5
Parte seconda	
Item 1	0,1
Item 2	0,7
Item 3	0,1
Item 4	0,4
Item 5	0,4

Tabella n. 4.13 - Indice di difficoltà degli *item* della quinta prova di ascolto

I valori IF che rientrano nella norma sono 5/10, dunque il 50% degli *item* presenta una buona percentuale di difficoltà, il 20% risulta troppo facile e il restante 30 % risulta troppo difficile. L'*item* n. 3 della prima parte è l'unico a cui tutti i candidati, di entrambi i gruppi, hanno risposto correttamente.

Prova di lettura

Prova n. 1	IF valore accettabile 0,3 - 0,7	Prova n. 2	IF valore accettabile 0,3 - 0,7	Prova n.3	IF valore accettabile 0,3 - 0,7	Prova n. 4	IF valore accettabile 0,3 - 0,7
Parte prima							
Item 1	1	Item 1	0,61	Item 1	0,67	Item 1	0,39
Item 2	1	Item 2	0,61	Item 2	0,45	Item 2	0,67
Item 3	0,9	Item 3	0,39	Item 3	1	Item 3	0,67
Item 4	0,8	Item 4	0,78	Item 4	1	Item 4	0,67
Item 5	0,95	Item 5	0,67	Item 5	0,61	Item 5	0,94
Parte seconda							
Item 1	0,85	Item 1	0,83	Item 1	0,67	Item 1	0,83
Item 2	0,95	Item 2	0,72	Item 2	0,78	Item 2	0,94
Item 3	0,8	Item 3	0,5	Item 3	0,83	Item 3	1
Item 4	0,45	Item 4	0,67	Item 4	0,5	Item 4	0,94
Item 5	0,7	Item 5	0,23	Item 5	0,23	Item 5	0,78
Parte terza							
Item 1	0,9	Item 1	0,61	Item 1	0,17	Item 1	0,56
Item 2	0,6	Item 2	0,5	Item 2	0,72	Item 2	0,83
Item 3	0,75	Item 3	0,45	Item 3	0,94	Item 3	0,94
Item 4	0,8	Item 4	0,83	Item 4	0,67	Item 4	0,94
Item 5	0,7	Item 5	0,5	Item 5	0,94	Item 5	0,89

Parte quarta							
Item 1	0,8	Item 1	0	Item 1	0,06	Item 1	0,72
Item 2	0,9	Item 2	0,72	Item 2	0,5	Item 2	0,61
Item 3	0,8	Item 3	0,06	Item 3	0,94	Item 3	0,06
Item 4	0,5	Item 4	0,39	Item 4	0,67	Item 4	0,78
Item 5	0,95	Item 5	0,67	Item 5	0,83	Item 5	0,78
Item 6	0,85	Item 6	0,5	Item 6	0,89	Item 6	0,94
Item 7	0,45	Item 7	0,56	Item 7	0,61	Item 7	0,67
Item 8	0,7	Item 8	0,68	Item 8	0,83	Item 8	0,56
Item 9	0,45	Item 9	0,56	Item 9	0,61	Item 9	0,67
Item 10	0,6	Item 10	0,34	Item 10	0,83	Item 10	0,67

Tabella n. 4.14 – Indice di difficoltà degli *item* delle prime quattro prove di lettura

Per le prime quattro prove di lettura, i valori IF che rientrano nella norma, contrassegnati nella tabella dal colore rosso, sono 46/100. Del restante, il 47% degli *item* risulta troppo facile poiché i valori sono molto vicini al +1,00 e il 7% troppo difficile poiché vicino allo 0.

I risultati della prova di lettura mostrano, in generale, un buon indice di difficoltà degli *item*, tuttavia risulta ancora presente un numero elevato di *item* troppo semplice anche per i candidati meno competenti. Ad esempio nella prima prova di lettura, entrambi i gruppi hanno risposto correttamente ai primi due *item* della prima parte del test:

Tipo di prova: lettura		Gruppo: A		Totale candidati: 20		Data: 31 marzo 2009	
N. <i>item</i>	Risp.	Candidati gruppo A _n : 10	p	Candidati gruppo A _i : 10	p	IF	
Parte prima							
1	a	10	1	10	1	1	
	b	-	0	-	0	0	
	c	-	0	-	0	0	
	O	-	0	-	0	0	
2	a	-	0	-	0	0	
	b	10	1	10	1	1	
	c	-	0	-	0	0	
	O	-	0	-	0	0	

Tabella n. 4.15 - Prova di lettura n. 1 Parte I *Item* n. 1 e n. 2

La quinta prova presenta un indice di difficoltà migliore rispetto alle prove della prima sessione, poiché, come mostrato nella tabella sottostante, 17 *item* di 25 rientrano nei valori compresi tra 0,3 e 0,7 ossia il 68%; il restante è equamente suddiviso tra i valori inferiori a 0,3 e superiori a 0,7, rispettivamente 4/25 e 4/25.

Prova n. 1	IF valore accettabile 0,3 - 0,7	Prova n. 2	IF valore accettabile 0,3 - 0,7	Prova n.3	IF valore accettabile 0,3 - 0,7	Prova n. 4	IF valore accettabile 0,3 - 0,7
Parte prima							
Item 1	1	Item 1	0,61	Item 1	0,67	Item 1	0,39
Item 2	1	Item 2	0,61	Item 2	0,45	Item 2	0,67
Item 3	0,9	Item 3	0,39	Item 3	1	Item 3	0,67
Item 4	0,8	Item 4	0,78	Item 4	1	Item 4	0,67
Item 5	0,95	Item 5	0,67	Item 5	0,61	Item 5	0,94
Parte seconda							
Item 1	0,85	Item 1	0,83	Item 1	0,67	Item 1	0,83
Item 2	0,95	Item 2	0,72	Item 2	0,78	Item 2	0,94
Item 3	0,8	Item 3	0,5	Item 3	0,83	Item 3	1
Item 4	0,45	Item 4	0,67	Item 4	0,5	Item 4	0,94
Item 5	0,7	Item 5	0,23	Item 5	0,23	Item 5	0,78
Parte terza							
Item 1	0,9	Item 1	0,61	Item 1	0,17	Item 1	0,56
Item 2	0,6	Item 2	0,5	Item 2	0,72	Item 2	0,83
Item 3	0,75	Item 3	0,45	Item 3	0,94	Item 3	0,94
Item 4	0,8	Item 4	0,83	Item 4	0,67	Item 4	0,94
Item 5	0,7	Item 5	0,5	Item 5	0,94	Item 5	0,89
Parte quarta							
Item 1	0,8	Item 1	0	Item 1	0,06	Item 1	0,72
Item 2	0,9	Item 2	0,72	Item 2	0,5	Item 2	0,61
Item 3	0,8	Item 3	0,06	Item 3	0,94	Item 3	0,06
Item 4	0,5	Item 4	0,39	Item 4	0,67	Item 4	0,78
Item 5	0,95	Item 5	0,67	Item 5	0,83	Item 5	0,78
Item 6	0,85	Item 6	0,5	Item 6	0,89	Item 6	0,94
Item 7	0,45	Item 7	0,56	Item 7	0,61	Item 7	0,67
Item 8	0,7	Item 8	0,68	Item 8	0,83	Item 8	0,56
Item 9	0,45	Item 9	0,56	Item 9	0,61	Item 9	0,67
Item 10	0,6	Item 10	0,34	Item 10	0,83	Item 10	0,67

Tabella n. 4.16 - Indice di difficoltà degli *item* della quinta prova di lettura

Nella prova di lettura non vi è alcun *item* con IF pari a 1, dunque non si è mai presentata la situazione in cui tutti i candidati hanno risposto correttamente. Al contrario, ad esempio, l'*item* n. 1 della parte seconda presenta la condizione opposta, con un IF pari a 0:

Tipo di prova: lettura		Gruppo: E		Totale candidati: 10		Data: 29 giugno 2009	
N. <i>item</i>	Risp.	Candidati gruppo E _n : 5	p	Candidati gruppo E _i : 5	p	IF	
Parte seconda							
1	a	1	0,2	-	0	0,1	
	b	4	0,8	5	1	0,9	
	c	-	0	-	0	0	
	O	-	0	-	0	0	

Tabella n. 4.17 - Prova di lettura n. 5 Parte II *Item* n. 1

Ciò equivale a una situazione in cui nessun candidato ha scelto la risposta corretta, neppure tra i candidati dei gruppi *h*.

Indice di distrattività

Per il calcolo dell'indice di distrattività è stata impiegata la formula n. 3.28

$$[p_x = \frac{p_h + p_l}{2}] \text{ oppure l'equivalente n. 3.27 } [p_x = \frac{nr_h + nr_l}{2n}].$$

Prova di ascolto

Gli indici di distrattività che rientrano nella norma, per le prime quattro prove di ascolto (per il quadro di tutte le risposte per ogni *item* di ciascuna prova si rimanda all'appendice), sono 5, così distribuiti:

- prova prima:
 - parte seconda, *item* 2, opzione *a* (0,3);
 - parte seconda, *item* 4, opzione *a* (0,85);
 - parte seconda, *item* 5, opzione *b* (0,45).
- prova seconda:
 - parte seconda, *item* 5, opzione *c* (0,39).
- prova terza:
 - nessuno.
- prova quarta:
 - parte seconda, *item* 5, opzione *c* (0,34).

In sintesi, su un totale di 80 distrattori (20 per ciascuna prova), solo il 6% di essi svolge la funzione di deviare il candidato dalla risposta corretta.

L'*item* n. 1 della prima prova di ascolto, parte seconda (tab. n. 4.12), presenta 2 distrattori con un indice pari a 0 per entrambi, dunque nessuno dei due svolge la propria funzione.

L'*item* n. 1 della terza prova di ascolto, parte seconda (tab. n. 4.18), ha un IF pari a 0,89 e due indici di distrattività inferiori alla soglia minima:

Tipo di prova: ascolto		Gruppo: C		Totale candidati: 18		Data: 31 marzo 2009	
N. <i>item</i>	Risp.	Candidati gruppo C _h : 9	p	Candidati gruppo C _i : 9	p	IF	
Parte seconda							
1	a	-	0	1	0,12	0,06	
	b	9	1	7	0,78	0,89	
	c	-	0	1	0,12	0,06	
	O	-	0	-	0	0	

Tabella n. 4.18 - Prova di ascolto n. 3 Parte II *Item* n. 1

È interessante notare che tutti i candidati del gruppo C_h hanno risposto correttamente a tale *item*; il gruppo C_i , al contrario, è stato distratto dalle opzioni *a* e *c* in maniera quasi irrilevante.

Nella quinta prova (l'intero prospetto in appendice), le opzioni che raggiungono un indice di distrattività superiore a 0,3 sono 9/20, ossia il 45%, e del restante 55% oltre la metà sono comunque uguali o maggiori di 0,1 (paragrafo 3.3.2).

Un buon esempio è l'*item* n. 3 della parte seconda, in cui i due distrattori hanno svolto efficacemente la loro funzione, distogliendo dalla risposta esatta anche candidati del gruppo *h*:

Tipo di prova: ascolto		Gruppo: E		Totale candidati: 10		Data: 29 giugno 2009	
N. <i>item</i>	Risp.	Candidati gruppo E_h : 10	p	Candidati gruppo E_i : 10	p	IF	
Parte seconda							
3	a	1	0,2	3	0,6	0,4	
	b	3	1,6	2	0,4	0,5	
	c	1	0,2	0	0	0,1	
	O	-	0	-	0	0	

Tabella n. 4.19 - Prova di ascolto n. 5 Parte II *Item* n. 3

Prova di lettura

Per le prime quattro prove di lettura (per il quadro di tutte le risposte per ogni *item* di ciascuna prova si rimanda all'appendice), gli indici di distrattività che rientrano nella norma sono 27 di 180 (45 per ciascuna prova), quindi il 15%.

Ad esempio l'*item* n. 2 della prima prova (parte quarta) presenta un IF pari a 0,9 e un indice di distrattività per entrambe le opzioni *a* e *c* pari a 0,05:

Tipo di prova: lettura		Gruppo: A		Totale candidati: 20		Data: 31 marzo 2009	
N. <i>item</i>	Risp.	Candidati gruppo A_h : 10	p	Candidati gruppo A_i : 10	p	IF	
Parte quarta							
2	a	-	0	1	0,1	0,05	
	b	10	1	8	0,8	0,9	
	c	-	0	1	0,1	0,05	
	O	-	0	0	0	0	

Tabella n. 4.20 - Prova di lettura n. 1 Parte IV *Item* n. 2

Nella quinta prova la percentuale di distrattori funzionali è aumentata fino

al 51% (23/45). Se si osserva, ad esempio, l'item n. 2 della seconda parte, si possono notare due distrattori ben funzionanti:

Tipo di prova: lettura		Gruppo: E		Totale candidati: 10		Data: 29 giugno 2009	
N. item	Risp.	Candidati gruppo E _n : 5	p	Candidati gruppo E _i : 5	p	IF	
Parte seconda							
2	a	1	0,2	4	0,8	0,5	
	b	3	0,6	1	0,2	0,4	
	c	1	0,2	-	0	0,1	
	O	-	0	0	0	0	

Tabella n. 4.21 - Prova di lettura n. 5 Parte II Item n. 2

Indice di discriminazione

Per calcolare l'indice di discriminazione sono state utilizzate le formule n.

3.33 $[ID = (p_h - p_l)]$ oppure n. 3.32 $[ID = \frac{R_h - R_l}{N}]$; per il calcolo dell'indice di discriminazione dei distrattori è stata impiegata la formula n. 3.35 $[ID = (p_h - p_l)]$ oppure l'equivalente n. 3.34 $[p_x = \frac{nr_h - nr_l}{n}]$.

Prova di ascolto

Prova n. 1	ID valore accettabile 0,3 - 0,7	Prova n. 2	ID valore accettabile 0,3 - 0,7	Prova n.3	ID valore accettabile 0,3 - 0,7	Prova n. 4	ID valore accettabile 0,3 - 0,7
Parte prima							
Item 1	0,1	Item 1	0,34	Item 1	0,23	Item 1	0,23
Item 2	0	Item 2	0,34	Item 2	0	Item 2	0,23
Item 3	0	Item 3	0	Item 3	0,67	Item 3	0
Item 4	0,6	Item 4	0,34	Item 4	0,34	Item 4	0
Item 5	0	Item 5	0,45	Item 5	0	Item 5	0
Parte seconda							
Item 1	0	Item 1	0,12	Item 1	0,23	Item 1	0,23
Item 2	0,8	Item 2	0	Item 2	0	Item 2	0,34
Item 3	0	Item 3	0,34	Item 3	0,12	Item 3	0
Item 4	0,3	Item 4	0,67	Item 4	0,23	Item 4	0,12
Item 5	0,3	Item 5	0,89	Item 5	0	Item 5	0,78

Tabella n. 4.22 – Indice di discriminazione degli item delle prime quattro prove di ascolto

Per le prime quattro prove di ascolto, i valori ID che rientrano nella norma, contrassegnati nella tabella dal colore rosso, sono 13/40, quindi il 32,5% degli item ha la capacità di stabilire una distinzione tra i candidati che hanno avuto una

migliore *performance* e quelli che sono stati meno competenti. Inoltre i valori ID sono positivi, dunque la risposta corretta è stata scelta da un numero maggiore di candidati appartenenti ai gruppi *h*.

Al contrario, per l'indice di discriminazione dei distrattori, 33 *item* su 80 (per il quadro di tutte le risposte per ogni *item* di ciascuna prova si rimanda all'appendice), presentano un valore negativo, ciò vuol dire che, per il 41% degli *item*, un numero maggiore di candidati appartenenti ai gruppi *l*, rispetto ai gruppi *h*, è stato attratto da uno dei distrattori.

Se si considera l'*item* n. 5 della prima prova di ascolto (parte seconda), si può notare un ID pari a 0,3, dunque hanno risposto correttamente i candidati migliori e una parte dei meno competenti:

Tipo di prova: ascolto		Gruppo: A		Totale candidati: 20		Data: 31 marzo 2009	
N. <i>item</i>	Risp.	Candidati gruppo A _h : 10	p	Candidati gruppo A _l : 10	p	ID	
Parte seconda							
5	a	-	0	-	0	0	
	b	3	0,3	6	0,6	-0,3	
	c	7	0,7	4	0,4	0,3	
	O	-	0	-	0	0	

Tabella n. 4.23 - Prova di ascolto n. 1 Parte II *Item* n. 5

Inoltre si può notare che l'indice di discriminazione del distrattore *b* è un valore negativo, di conseguenza il distrattore *b* svolge la propria funzione in modo efficiente ma non è in grado di discriminare notevolmente tra il gruppo *A_h* e il gruppo *A_l*.

Nella quinta prova, come si evince dalla tabella sottostante, 4 *item* di 10 (40%) hanno raggiunto un indice di discriminazione compreso tra i valori accettabili; inoltre gli *item* n. 5 della prima e seconda parte hanno una discriminazione massima pari rispettivamente a 1 e 0,8.

Prova n. 5	ID valore accettabile 0,3 - 0,7
Parte prima	
Item 1	0,4
Item 2	0,6
Item 3	0
Item 4	-0,4
Item 5	1
Parte seconda	
Item 1	0,2
Item 2	0,2
Item 3	0,2
Item 4	0,4
Item 5	0,8

Tabella n. 4.24 - Indice di discriminazione degli *item* della quinta prova di ascolto

Nella quinta prova (l'intera tabella in appendice), le opzioni che presentano un indice di discriminazione negativo, ossia una capacità di discriminazione tra gli allievi più preparati e quelli meno preparati, sono 10/20, ossia il 50%.

L'*item* n. 2 della prima parte della quinta prova mostra un ID perfetto e due indici di discriminazione dei distrattori in grado di discriminare considerevolmente tra il gruppo *E_h* e il gruppo *E_l*:

Tipo di prova: ascolto		Gruppo: E		Totale candidati: 10		Data: 29 giugno 2009	
N. item	Risp.	Candidati gruppo E _h : 5	p	Candidati gruppo E _l : 5	p	ID	
Parte prima							
2	a	-	0	1	0,2	-0,2	
	b	1	0,2	3	0,6	-0,4	
	c	4	0,8	1	0,2	0,6	
	O	-	0	-	0	0	

Tabella n. 4.25 - Prova di ascolto n. 5 Parte I *Item* n. 2

Prova di lettura

Prova n. 1	ID valore accettabile 0,3 - 0,7	Prova n. 2	ID valore accettabile 0,3 - 0,7	Prova n.3	ID valore accettabile 0,3 - 0,7	Prova n. 4	ID valore accettabile 0,3 - 0,7
Parte prima							
Item 1	0	Item 1	0,11	Item 1	0,66	Item 1	0,78
Item 2	0	Item 2	0,55	Item 2	0,44	Item 2	0,22
Item 3	0,2	Item 3	0,55	Item 3	0	Item 3	0,66
Item 4	0,4	Item 4	0	Item 4	0	Item 4	0,44
Item 5	0,1	Item 5	0,44	Item 5	0,55	Item 5	0,11

Parte seconda							
Item 1	0,1	Item 1	0,33	Item 1	0,44	Item 1	0,33
Item 2	0,1	Item 2	0,55	Item 2	0,44	Item 2	0,11
Item 3	0,4	Item 3	0,55	Item 3	0,33	Item 3	0
Item 4	0,9	Item 4	0,66	Item 4	0,33	Item 4	-0,11
Item 5	0,6	Item 5	0	Item 5	0,22	Item 5	0,44
Parte terza							
Item 1	0,2	Item 1	0,77	Item 1	-0,11	Item 1	0,66
Item 2	0	Item 2	0,55	Item 2	0,55	Item 2	0,33
Item 3	-0,1	Item 3	0,66	Item 3	0,11	Item 3	0,11
Item 4	0,2	Item 4	0,33	Item 4	0,22	Item 4	0,11
Item 5	0,4	Item 5	0,33	Item 5	-0,11	Item 5	0,22
Parte quarta							
Item 1	0,4	Item 1	0	Item 1	0,12	Item 1	0,11
Item 2	0,2	Item 2	0,55	Item 2	0,77	Item 2	0,77
Item 3	0	Item 3	-0,12	Item 3	0,11	Item 3	-0,12
Item 4	0,2	Item 4	0,33	Item 4	-0,22	Item 4	0,44
Item 5	0,1	Item 5	0,66	Item 5	0,33	Item 5	0,44
Item 6	0,3	Item 6	0,77	Item 6	0,22	Item 6	0,11
Item 7	0,7	Item 7	0,66	Item 7	0,11	Item 7	0,66
Item 8	0,6	Item 8	0,66	Item 8	0,33	Item 8	0,66
Item 9	0,1	Item 9	0,22	Item 9	0,77	Item 9	0,66
Item 10	0,8	Item 10	0	Item 10	0,33	Item 10	0,66

Tabella n. 4.26 – Indice di discriminazione degli *item* delle prime quattro prove di lettura

Per le prime quattro prove di lettura, i valori ID che rientrano nella norma, contrassegnati nella tabella dal colore rosso, sono 47/100. Purtroppo, sebbene una buona percentuale di *item* (il 47%) possieda un valido potere discriminante tra i candidati più competenti e quelli meno competenti, sono presenti alcuni valori negativi tra gli ID (evidenziati in viola), pertanto, in alcuni *item*, un numero maggiore di candidati dei gruppi *l* ha scelto la risposta corretta rispetto ai gruppi *h*.

Gli ID negativi risultano in quantità elevata anche tra i distrattori (lo schema completo in appendice), per un totale di 102/180, per cui il 57% degli *item* contiene distrattori in grado di attrarre un numero considerevole di candidati appartenenti ai gruppi *l*.

Se, ad esempio, si prendono in esame i valori dell'*item* n. 2 della prima parte della quarta prova:

Tipo di prova: lettura		Gruppo: D		Totale candidati: 18		Data: 31 marzo 2009	
N. item	Risp.	Candidati gruppo Dh: 9	p	Candidati gruppo Dl: 9	p	ID	
Parte prima							
2	a	2	0,23	4	0,45	-0,22	
	b	7	0,78	5	0,56	0,22	
	c	-	0	-	0	0	
	O	-	0	-	0	0	

Tabella n. 4.27 - Prova di lettura n. 4 Parte I Item n. 2

ha risposto correttamente una buona percentuale di candidati del gruppo Dh e anche una parte di candidati del gruppo Dl, inoltre il distrattore a ha svolto la propria funzione attirando candidati da entrambi i gruppi, ma non è stato in grado di discriminare molto tra il gruppo Dh e il gruppo Dl.

Nella quinta prova (tab. n. 4.29) 16 item di 25 (64%) presentano un indice di discriminazione compreso tra i valori accettabili:

Prova n. 5	ID valore accettabile 0,3 - 0,7	Prova n. 5	ID valore accettabile 0,3 - 0,7	Prova n. 5	ID valore accettabile 0,3 - 0,7
Parte prima		Parte terza		Parte quarta	
Item 1	0,6	Item 1	0,6	Item 1	0,2
Item 2	0,6	Item 2	0,2	Item 2	0,4
Item 3	0,4	Item 3	0,6	Item 3	0,4
Item 4	0	Item 4	0,4	Item 4	0,6
Item 5	0,6	Item 5	0,4	Item 5	1
Parte seconda				Item 6	0,4
Item 1	0			Item 7	1
Item 2	0,2			Item 8	0,8
Item 3	0,4			Item 9	0,8
Item 4	0,6			Item 10	0,6
Item 5	0,4				

Tabella n. 4.28 - Indice di discriminazione degli item della quinta prova di lettura

Nella quinta prova (l'intero quadro in appendice), le opzioni che presentano un indice di discriminazione negativo, ossia una capacità di discriminazione tra i candidati competenti e i meno preparati, costituiscono il 58% (26/45).

Se si considera, ad esempio, l'item n. 5 della prima parte della quinta prova, si nota un ID nella norma e due indici di discriminazione dei distrattori a e b capaci di distinguere i candidati competenti dai non:

Tipo di prova: lettura		Gruppo: E		Totale candidati: 10		Data: 29 giugno 2009	
N. item	Risp.	Candidati gruppo E _h : 5	p	Candidati gruppo E _l : 5	p	ID	
Parte prima							
6	a	-	0	2	0,4	-0,4	
	b	-	0	1	0,2	-0,2	
	c	5	1	2	0,4	0,6	
	O	-	0	-	0	0	

Tabella n. 4.29 - Prova di lettura n. 5 Parte I Item n. 5

4.3.2 Analisi linguistica degli *item*

Analizzando, da un punto di vista linguistico, gli *item* presentati nel paragrafo precedente come esempi validi o inefficaci di IF o ID, si possono comprendere, sulla base degli indicatori statistici, gli aspetti che rendono un candidato più o meno abile nella sua competenza linguistico-comunicativa.

L'*item* n. 1 della seconda parte della prima prova di ascolto (tab. n. 4.13) appare, attraverso l'ascolto del testo oggetto di prova della comprensione orale, eccessivamente semplice per via dell'attenzione che il parlante madrelingua pone sull'espressione: "... for two days, not two weeks, ...", fornendo in tal modo la risposta corretta (b), mediante l'utilizzo degli stessi termini presenti nella stessa. L'*item* si presenta infatti in tal modo:

How long will the trip last?
a) On day.
b) <u>Two days.</u>
c) A week.

Una situazione simile si può riscontrare nell'*item* n. 2 della seconda parte della prima prova di ascolto (tabella in appendice):

2. The night before she had...
a) dinner.
b) just a sandwich.
c) <u>nothing to eat at all.</u>

in cui tutti i candidati del gruppo *h* (10/10) e 9/10 del gruppo *l* hanno scelto

l'opzione corretta dopo aver letto il testo che riporta le seguenti parole: "... *she hadn't eaten anything the night before, ...*".

La situazione sembra essere del tutto diversa nella quinta prova, la quale presenta l'*item* n. 1 della seconda parte della prova di lettura (tab. n. 4.17) sotto questa veste:

1. The Rome-Milan "Motorway of the Sun"...
a) makes everyone go to sleep.
b) is very straight and flat.
c) <u>is wide and has a level surface.</u>

Il primo paragrafo del testo di lettura di riferimento sul quale si basa l'*item* recita: "*The first day Mr Baxter got to Milan on the Rome-Milan "motorway" as the English call it. It is wonderful! It's broad and smooth but not so straight and flat as the motorway from Milan to Turin. Mr Baxter went along the Milan-Turin road on the second day, and it's so flat and straight that some people go to sleep. But luckily Mr Baxter didn't! The beauty of the Rome-Milan "Motorway of the Sun" consists in the fact that it curves up and down through the mountains, the Apennines, especially between Bologna and Florence. Despite the curves, Mr Baxter drove fast all the way.*"

È evidente che la complessità nella risposta esatta consiste nella conoscenza dei termini "*broad*" e "*smooth*" presenti nel testo e dei sinonimi corrispondenti "*wide*" e "*level*" presenti nella risposta corretta. Inoltre la complessità del paragrafo a livello perifrastico influisce sulla difficoltà dell'*item*. La complessità del testo in ogni caso appare adeguata a quanto esposto nel QCER a proposito delle capacità di ricezione scritta relative al livello B1: "*È in grado di scorrere testi di una certa lunghezza alla ricerca di informazioni specifiche e di reperire in punti diversi di un testo o in testi diversi le informazioni necessarie per portare a termine un compito specifico*" (QCER, 2001:88).

Un *item* che mostra un adeguato IF (tabella in appendice) è il n. 5 della prima parte della quinta prova di ascolto:

5. How old is the man?
a) <u>49</u> .
b) 50.
c) 15.

poiché il dialogo di riferimento tra due parlanti si presenta come segue:

A: *Happy birthday!*

B: *My birthday is next Sunday.*

A: *Oh, dear. Sorry. So you are not 50 yet.*

B: *Not until Sunday.*

È facilmente comprensibile che, nonostante la facilità del dialogo e la presenza del numero 50 sia nel testo che in una delle opzioni di risposta, l'*item* presenta una trappola, dunque solo i candidati che possiedono realmente una competenza linguistica adeguata al compito riescono a comprendere il dialogo e a rispondere correttamente.

Il QCER afferma, a proposito della capacità di ricezione orale relativa al livello B1, che il candidato “è *in grado di comprendere informazioni fattuali chiare su argomenti comuni relativi alla vita di tutti i giorni o al lavoro, riconoscendo sia il significato generale sia le informazioni specifiche...*” (QCER, 2001:83).

Esaminando l'indice di distrattività dell'*item* n. 1 della seconda parte della terza prova di ascolto (tab. n. 4.18):

1. The old lady was standing...
a) in front of the bank.
b) <u>in the queue.</u>
c) next to the counter.

si può notare l'inadeguatezza dei distrattori, i quali contengono termini che non appaiono mai nel racconto del parlante, e ancora una volta la risposta esatta

contiene la stessa espressione pronunciata dalla voce narrante: “*I was standing in the queue, waiting to pay my pension...*”.

Una condizione totalmente differente è rintracciabile all'interno dell'*item* n. 3 della seconda parte della quinta prova di ascolto (tab. n. 4.19), in cui i distrattori contengono tutti la stessa parola chiave “*sports*”:

3. The Martin family...
a) are good at sports.
b) like sports more than the Rossi family.
c) <u>are keen on sports.</u>

cosicché il candidato deve mostrare un'accurata comprensione del testo (il quale recita: “*The Martins are very fond of sports...*”) e delle possibili risposte, e deve sapere o riuscire a intuire che “*keen*” e “*fond*” sono sinonimi tra loro.

Se si prende in esame l'ID, paragonando due *item* della stessa tipologia appartenenti a prove di due diverse sessioni d'esame, si può notare una considerevole capacità di discriminazione nell'*item* della seconda somministrazione, dovuta a una migliore formulazione dei distrattori:

Prova I (I somministrazione)	Prova V (II somministrazione)
5. What's going to happen in the evening?	5. David...
a) A disco and then a meal.	a) plays sports with his brothers and sister.
b) A walk and then a disco.	b) <u>enjoys himself in the garden with his brothers and sister.</u>
c) <u>A meal and then a disco.</u>	c) plays with his toys in the house.

All'interno dell'*item* n. 5 della prima parte della prova di ascolto (tab. n. 4.23) le opzioni presentano tutte la parola chiave “disco”, ciò vuol dire che solo un ascoltatore attento, capace di cogliere il secondo termine e l'ordine degli elementi, potrà scegliere la risposta corretta (il testo di riferimento pronunciato dal parlante è il seguente: “*There will be dinner for the whole group, followed by dancing at a disco.*”). L'ID di questo *item* è pari a 0,3; 3/10 candidati del gruppo *h*

e 6/10 del gruppo I hanno risposto in modo errato, rispettivamente dunque 7/10 e 4/10 hanno risposto correttamente.

Attraverso l'*item* n. 5 della seconda parte della quinta prova di ascolto (tabella in appendice) si è puntato verso una comprensione più profonda del testo (le parole di riferimento pronunciate dal parlante sono le seguenti: "*While the elder children practise sports, David the baby, plays with his toys on the grass.*"), infatti le tre opzioni non mostrano parole chiave identiche, ma tre possibili risposte, delle quali la prima è vera per altri soggetti (*the elder children*), la terza può apparire parzialmente vera solo se ci si sofferma su "*toys*" e può attrarre soltanto se non si conosce il termine "*grass*" presente nel testo. Come mostra l'indagine statistica l'ID di tale *item* è pari a 0,8, 4/5 candidati del gruppo *h* e 0/10 del gruppo *I* hanno risposto in modo esatto, rispettivamente 1/5 e 5/5 hanno risposto erroneamente (2 hanno scelto l'opzione *a* e 3 l'opzione *c*), poiché probabilmente il gruppo *I* è stato attratto dal verbo "*plays*" presente sia nel testo che nella risposta, oppure dal termine "*toys*", o ancora da "*sports*". L'opzione *b*, ossia la risposta corretta, è l'unica che non presenta parole uguali al testo di riferimento da cui si deve solo desumere il significato generale e associarlo a quello pronunciato dal parlante.

La stessa situazione è ravvisabile nell'*item* n. 5 della prima parte della quinta prova di lettura (tab. n. 4.29):

5. Choose the right answer:
a) Janet says she can start work at any time of the day.
b) Janet says she can't have an interview with Mr Thomas at any time.
c) <u>Janet says she is waiting for Mr Thomas to contact her.</u>

in cui la risposta esatta presenta parole diverse rispetto al testo: "*I can be available for interview at any time. I look forward to hearing from you.*" e ne va dedotto solo il significato.

Riepilogo e confronto

Attraverso gli indici calcolati in appendice è facilmente possibile risalire agli *item* che risultano troppo facili o ai distrattori poco funzionali, ed effettuare quindi le opportune modifiche ai contenuti.

I dati della quinta prova (II sessione), come mostra la tabella delle percentuali sottostante, sembrano supportare l'ipotesi che, dopo un'indagine statistica e una successiva elaborazione qualitativa del contenuto, i test risultano maggiormente adatti allo scopo per cui sono stati realizzati.

	IF	IF di P _x	ID	ID di P _x
Prove di ascolto I sessione	17,5%	6%	32,5%	41%
Prove di ascolto II sessione	50%	45%	40%	50%
Prove di lettura I sessione	46%	15%	47%	57%
Prove di lettura II sessione	68%	51%	64%	58%

Tabella n. 4.30 – Percentuali degli indici presi in esame

4.3.3 Analisi delle distribuzioni di punteggi

L'analisi delle distribuzioni dei punteggi delle cinque prove è stata condotta attraverso il calcolo della media (paragrafo 3.3.1) per verificare la tendenza centrale dei candidati e della deviazione standard (paragrafo 3.3.1) per la variabilità dei dati della distribuzione rispetto alla media generale.

L'analisi è basata sulle cinque prove sia di lettura che di ascolto, suddivise sempre in due blocchi sulla base della prima (marzo) e seconda (giugno) sessione di somministrazione delle prove.

Si è proceduto al calcolo della media per ciascuna prova per ognuno dei cinque gruppi attraverso la formula n. 3.5 [$M a = \frac{\sum x}{N}$] e/o n. 3.6 [$M a \text{ pond} = \frac{\sum fx}{N}$].

La media calcolata sui punteggi ottenuti nelle prove di ascolto (in cui il punteggio massimo raggiungibile è 20) è mostrata, per ciascuno dei cinque gruppi, nel grafico in basso (la tabella completa delle operazioni in appendice):

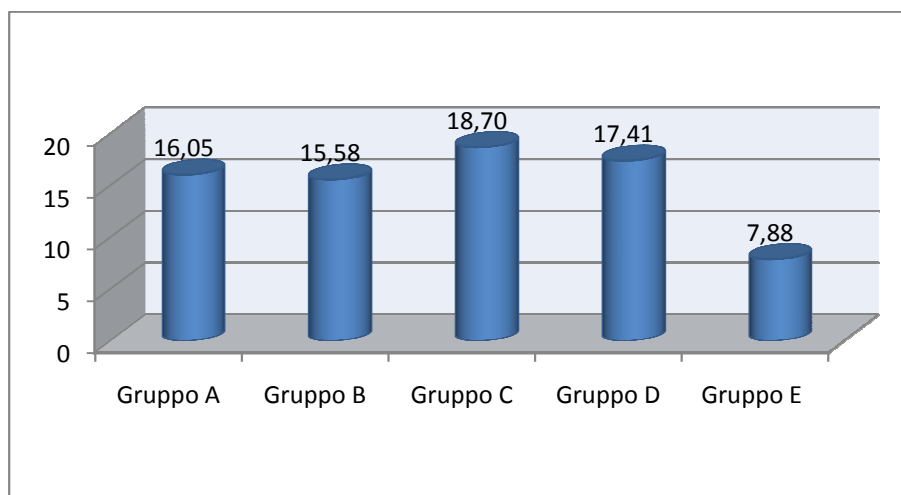


Figura n. 4.1- Media dei punteggi di ciascun gruppo per le prove di ascolto

Dalla rappresentazione grafica si evince che la media migliore è stata raggiunta dal gruppo C. È evidente che il gruppo con la media più bassa è il gruppo E, ossia il gruppo a cui è stata somministrata la quinta prova, la quale è risultata, dall'analisi degli *item*, la prova che ha raggiunto la più alta percentuale di IF o ID appartenenti alla norma.

La media calcolata sui punteggi ottenuti nelle prove di lettura (in cui il punteggio massimo raggiungibile è 25) è mostrata, per ciascun gruppo, nel grafico in basso:

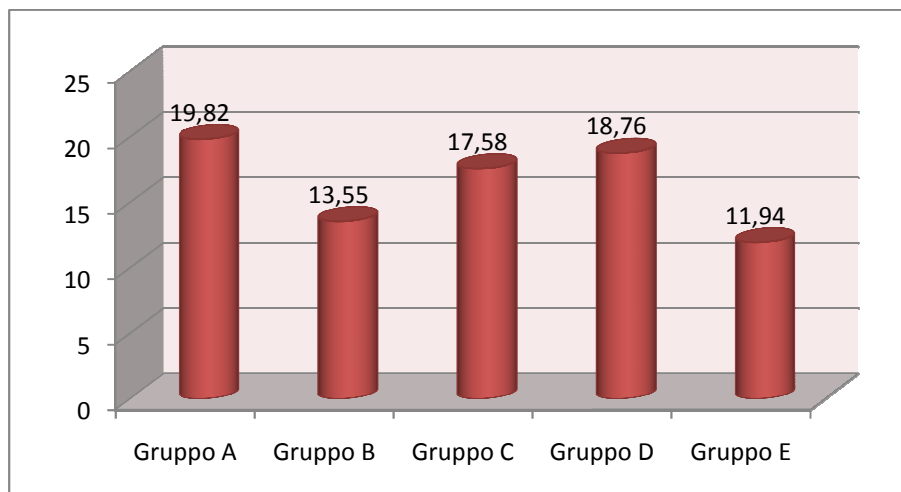


Figura n. 4.2 - Media dei punteggi di ciascun gruppo per le prove di lettura

Dal grafico si evince che la media migliore è stata raggiunta dal gruppo A. Anche in questo caso il gruppo E risulta essere quello con la media più bassa, poiché la quinta prova ha ottenuto la più alta percentuale di IF o ID appartenenti alla norma.

Per descrivere interamente una data distribuzione è necessario disporre, oltre che della media, di informazioni aggiuntive riguardanti la dispersione dei punteggi intorno alla media stessa. È possibile determinare, se si conosce la media di una distribuzione, qual è la percentuale di punteggi inferiori o superiori alla media stessa attraverso il calcolo della deviazione standard mediante la

formula n. 3.11 $[DS = \sqrt{\frac{\sum (x-M)^2}{N}}]$.

Sottraendo dalla media e sommando alla media stessa la deviazione standard, si ottengono rispettivamente i due punteggi soglia all'interno dei quali si colloca la percentuale di candidati che ha conseguito un punteggio prossimo alla media (secondo le percentuali calcolate in base alla curva di Gauss, in tale intervallo si dovrebbe collocare il 68% dei candidati). Al di sotto e al di sopra di tali soglie si dovrebbe collocare rispettivamente una percentuale di candidati che ha conseguito i punteggi più bassi e una con i punteggi più alti (la percentuale

dovrebbe, in una distribuzione normale, essere pari al 16% in entrambi i casi).

Prova di ascolto					
	Prima	Seconda	Terza	Quarta	Quinta
M	16,05	15,58	18,7	17,41	7,88
DS	1,88	3,17	1,81	2,25	2,32
M - DS	14,17	12,41	16,89	15,16	5,56
%	20%	20,5%	15%	15%	0%
M + DS	17,93	18,75	20,51	19,66	10,20
%	34%	20,5%	0%	26,5%	18%
M ± DS %	46%	59%	85%	58,5%	82%
Prova di lettura					
	Prima	Seconda	Terza	Quarta	Quinta
M	19,82	13,55	17,58	18,76	11,94
DS	2,67	3,90	3,03	3,46	5,10
M - DS	17,15	9,65	14,55	15,30	6,84
%	20%	20,5%	20,5%	26,5%	18%
M + DS	22,49	17,45	20,61	22,22	17,04
%	11%	23,5%	6%	18%	18%
M ± DS %	69%	56%	73,5%	55,5%	64%

Tabella n. 4.31 – Rappresentazione percentuale delle distribuzioni

Dalla rappresentazione grafica dei dati calcolati per le cinque prove (in appendice la tavola completa delle operazioni) emerge che, nella maggioranza dei casi, le percentuali dei punteggi che si collocano nell'intervallo $M \pm DS$ si avvicinano al 68% gaussiano.

La prova nella quale emerge il più alto tasso di dispersione dei dati dalla media centrale è la quinta prova di lettura, in cui la DS è rappresentata da un valore piuttosto alto (pari a 5,10), il quale indica un ampio raggio di distribuzione dei punteggi rispetto alla media.

La misura della tendenza centrale fornisce informazioni importanti anche se utilizzata per interpretare il punteggio di ogni singolo candidato.

Se, ad esempio, si volesse prendere in considerazione un punteggio, pari a 12, apparentemente identico, conseguito dal candidato n. 113 nella quarta prova di ascolto e dal candidato n. 138 nella quinta prova di ascolto (il prospetto completo in appendice), si noterebbe che, interpretato sulla base della distribuzione dei punteggi e della media della distribuzione, esso assumerebbe

un significato totalmente diverso a seconda dei casi: nel primo caso il candidato n. 113 ha conseguito il punteggio più basso in una distribuzione con una media pari a 17,41; nel secondo caso il candidato n. 138 ha ottenuto il punteggio più alto in una distribuzione con una media pari a 7,88. In conclusione un punteggio non possiede un valore universale, ma assume un valore relativo quando viene commisurato con gli altri punteggi e con la media della distribuzione.

4.3.4 Analisi statistica delle *performance*

L'analisi delle *performance* è stata condotta, per ciascuna delle cinque prove, attraverso il calcolo del coefficiente di correlazione tra le abilità di ascolto e di lettura, il quale mira a verificare se i livelli di abilità di ascolto e di lettura covariano, ossia se un certo livello di competenza sia stato raggiunto da un dato gruppo in entrambe le abilità.

Per verificare che le due abilità variano in maniera dipendente l'una dall'altra è necessario calcolare il coefficiente di correlazione mediante le formule n. 3.16 e n. 3.17 per il coefficiente di Pearson e la formula n. 3.18 per il coefficiente di Spearman (paragrafo 3.3.1).

Prima di correlare le due variabili si devono verificare alcune informazioni relativamente alle distribuzioni dei punteggi di ciascuna prova. Si necessita soprattutto di un controllo della forma di distribuzione dei punteggi e della scala di misurazione. Se le distribuzioni sono normali si può procedere al calcolo del coefficiente di correlazione mediante la formula sia di Pearson che di Spearman. Se le distribuzioni non sono normali si può procedere soltanto al calcolo del coefficiente tramite la formula di Spearman.

Per verificare che ogni distribuzione sia distribuita normalmente, si rappresentano graficamente, in un poligono delle frequenze, i punteggi sull'asse delle ascisse e le frequenze sull'asse delle ordinate, e si osserva se la forma di ciascuna distribuzione è simile alla campana di Gauss:

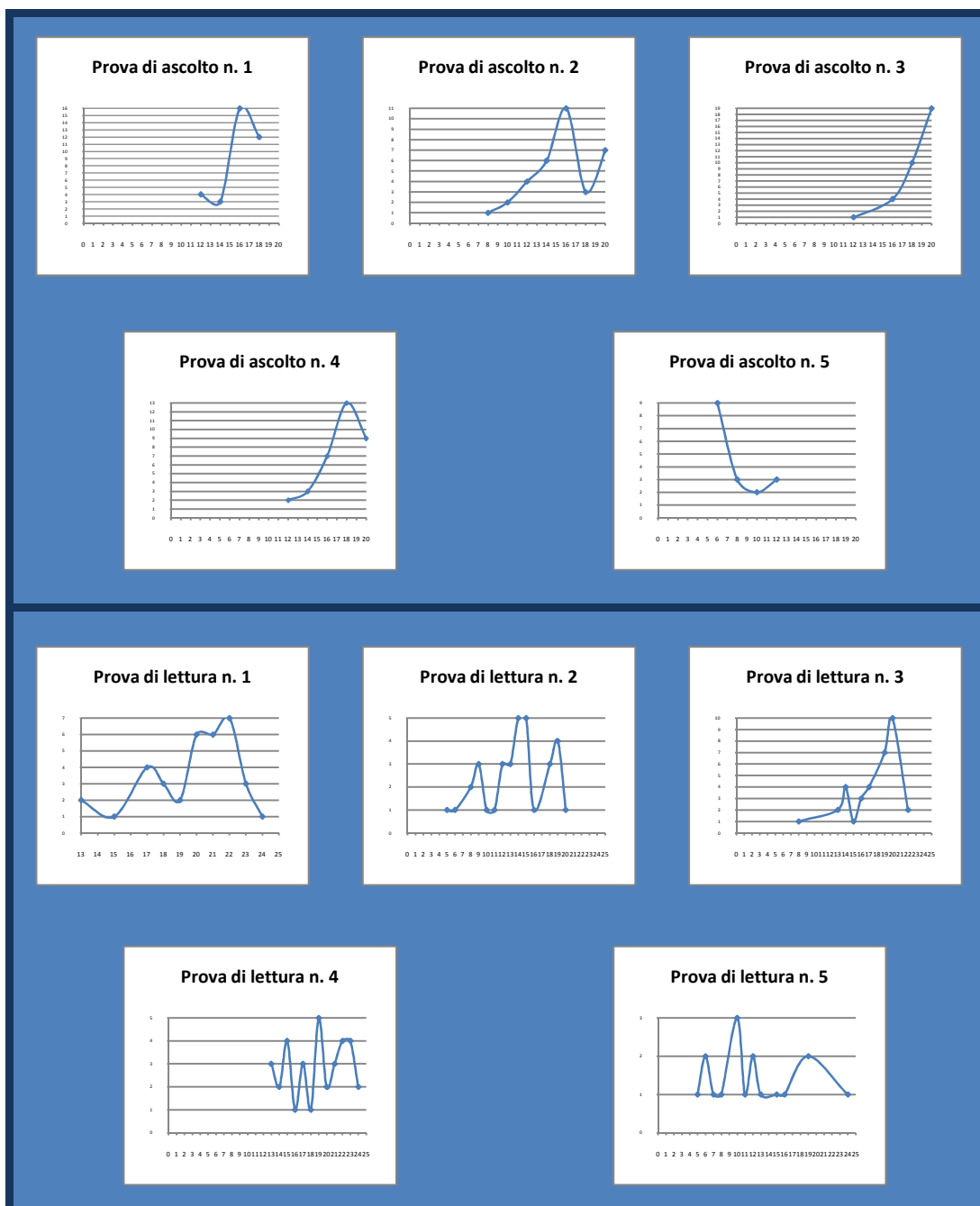


Tabella n. 4.32 – Poligono delle frequenze delle cinque distribuzioni di punteggi

Dall'osservazione dei grafici si può affermare che nessuna prova presenta una distribuzione normale, pertanto si può calcolare soltanto il coefficiente di

correlazione di Spearman $[\rho_s = 1 - \frac{6 \cdot \Sigma D^2}{N \cdot (N^2 - 1)}]$.

Per avere i punteggi su una stessa scala di misurazione è necessario trasformare i punteggi grezzi in punteggi standardizzati Z o T . Poiché le prove di ascolto e di lettura prese qui in esame sono basate su scale diverse (rispettivamente su un totale di 20 e di 25 punteggi), i punteggi ottenuti sono stati trasformati, mediante la formula n. 3.13 $[T = 50 + (10 \frac{(X-M)}{DS})]$, in punteggi T (in appendice la tabella delle operazioni al completo).

Successivamente, per calcolare il coefficiente di Spearman, il quale si basa sulle differenze di posizione delle variabili nelle due distribuzioni, si procede alla organizzazione dei punteggi (per ciascuna distribuzione) in una graduatoria in cui si attribuisce a ciascun punteggio una posizione in classifica (paragrafo 3.3.1). In seguito si confrontano le posizioni di ciascun candidato tra la graduatoria relativa alla prova di ascolto e quella relativa alla prova di lettura per ciascuna delle cinque prove, e si calcolano le differenze di posizione (lo schema completo delle operazioni in appendice).

A questo punto si può calcolare il coefficiente di correlazione di Spearman:

Prova n. 1
$\rho_s = 1 - \frac{6 \cdot 2489,5}{35 \cdot (35^2 - 1)} = 1 - \frac{14937}{35 \cdot (1225 - 1)} = 1 - \frac{16282,5}{42840} = 1 - 0,34 = 0,66$
Prova n. 2
$\rho_s = 1 - \frac{6 \cdot 5761}{34 \cdot (34^2 - 1)} = 1 - \frac{34566}{34 \cdot (1156 - 1)} = 1 - \frac{34566}{39270} = 1 - 0,88 = 0,12$
Prova n. 3
$\rho_s = 1 - \frac{6 \cdot 5010,5}{34 \cdot (34^2 - 1)} = 1 - \frac{30063}{34 \cdot (1156 - 1)} = 1 - \frac{30063}{39270} = 1 - 0,77 = 0,23$
Prova n. 4
$\rho_s = 1 - \frac{6 \cdot 3286,5}{34 \cdot (34^2 - 1)} = 1 - \frac{19719}{34 \cdot (1156 - 1)} = 1 - \frac{19719}{39270} = 1 - 0,50 = 0,50$

Prova n. 5

$$\rho_s = 1 - \frac{6 \cdot 647,5}{17 \cdot (17^2 - 1)} = 1 - \frac{3885}{17 \cdot (289 - 1)} = 1 - \frac{3885}{4896} = 1 - 0,79 = 0,21$$

Tabella n. 4.33 – Coefficiente di correlazione di Spearman

Il coefficiente di Spearman può variare tra +1,00 e -1,00: quanto più i dati ottenuti si avvicinano a tali valori tanto più la correlazione è forte; se, invece, i dati si avvicinano allo 0 la correlazione è debole. Secondo quanto affermato nel capitolo precedente un valore pari a 0,4 è convenzionalmente un coefficiente accettabile di correlazione positiva. Pertanto nel caso delle prove prese in esame, le prove n. 1 e n. 4 mostrano, nel complesso, una buona correlazione tra le abilità di ascolto e di lettura possedute dai candidati; le prove n. 2, n. 3 e n. 5 presentano una correlazione quasi assente.

Si riportano gli esempi di tre candidati che mostrano una correlazione tra le abilità di ascolto e di lettura su tre diversi livelli:

- il candidato n. 139 ha ottenuto un punteggio di 6/20 (equivalente al punteggio T di 41,91) nella prova di comprensione orale e di 8/25 (equivalente al punteggio T di 42,28) nella prova di comprensione scritta, mostrando un livello basso di competenza in entrambe le abilità;
- il candidato n. 6 ha ottenuto un punteggio di 16/20 (equivalente al punteggio T di 49,73) nella prova di comprensione orale e di 20/25 (equivalente al punteggio T di 50,67) nella prova di comprensione scritta, mostrando un livello medio-alto di competenza in entrambe le abilità;
- il candidato n. 1 ha ottenuto un punteggio di 18/20 (equivalente al punteggio T di 60,36) nella prova di comprensione orale e di 24/25 (equivalente al punteggio T di 65,67) nella prova di comprensione scritta, mostrando un livello alto di competenza in entrambe le abilità.

Questi tre esempi mostrano come i candidati, dopo aver frequentato i corsi di formazione linguistica ed essersi preparati all'esame finale, abbiano raggiunto una competenza linguistica (idonea o carente) quasi proporzionata tra l'abilità di lettura e l'abilità di ascolto.

Si riportano altresì due esempi di candidati che mostrano un'assenza di

correlazione tra le due abilità:

- il candidato n. 154 ha ottenuto un punteggio di 12/20 (equivalente al punteggio T di 67,73) nella prova di comprensione orale e di 24/25 (equivalente al punteggio T di 73,63) nella prova di comprensione scritta, mostrando un livello diverso di competenza nelle due abilità;
- il candidato n. 123 ha ottenuto un punteggio di 20/20 (equivalente al punteggio T di 61,51) nella prova di comprensione orale e di 13/25 (equivalente al punteggio T di 33,33) nella prova di comprensione scritta, mostrando un livello diverso di competenza nelle due abilità.

Questi ultimi due esempi mostrano come i candidati, dopo aver frequentato i corsi di formazione linguistica ed essersi preparati all'esame finale, abbiano raggiunto una competenza linguistica diversa nelle abilità di lettura e di ascolto, non adeguata dunque allo scopo.

Si potrebbe affermare, in conclusione, che i candidati (come quelli presi in considerazione come esempio) che hanno sostenuto le prove n. 2, n. 3 e n. 5 rientrano in una delle due seguenti ipotesi:

- reale possesso di livelli di competenza diversi nelle due abilità;
- possesso di livelli di competenza uguali nelle due abilità ma al conseguimento del punteggio più alto ha contribuito il parametro del guessing.

4.4 Riflessioni conclusive

Il fine di questo capitolo è, in primo luogo, quello di illustrare l'applicazione pratica di alcune formule presentate nel capitolo precedente, di mostrare la validità di tali strumenti attraverso un esempio tangibile e dunque di offrire uno strumento utile e non troppo artificioso a chi si occupa di *testing* in piccoli ambiti didattici. Qui sono state messe in pratica alcune formule (e altre sono state tralasciate), adatte ai test selezionati, tali da poter mostrare la vasta gamma di

informazioni ottenibili da un'analisi statistica. Dalla rilevazione presentata in questo capito si può notare che non sempre occorre applicare l'intero inventario delle formule statistiche. Se i dati sono piuttosto omogenei e i candidati non sono molto numerosi, non è necessario svolgere rilevazioni approfondite, spesso è sufficiente il calcolo della media, della deviazione standard¹⁷, della correlazione e l'analisi della distribuzione dei punteggi.

Al tempo stesso, in modo complementare, le formule impiegate hanno permesso di analizzare l'adeguatezza degli *item* dei test somministrati dal CLA di Palermo e di verificare alcuni aspetti delle *performance* dei candidati sottoposti a esame. Dall'analisi delle prove somministrate ai candidati e delle competenze linguistico-comunicative dei candidati, a distanza di quasi due anni dall'intervento di formazione, sono emerse informazioni utili sulle quali è stato possibile ed è tuttora possibile intervenire qualitativamente, sia dal punto di vista di un miglioramento dei sistemi valutativi e del contenuto delle prove, sia dal punto di vista di un miglioramento dell'offerta formativa.

¹⁷ Per ottimizzare i conteggi è possibile servirsi di tavole numeriche dei quadrati e delle radici quadrate, disponibili nella maggior parte dei testi di matematica e statistica.

Conclusioni generali

Dalla panoramica presentata nella prima parte di questo lavoro emerge una grande importanza attribuita alla conoscenza delle lingue in ambito europeo e oltre, e un interesse sempre maggiore verso l'apprendimento linguistico.

È evidente, di conseguenza, l'urgenza con cui si rende necessario concedere la stessa importanza ai sistemi di valutazione e certificazione delle lingue, e alla rivoluzione subita dai sistemi stessi in seguito al QCER e all'affermarsi di un linguaggio comune di organizzazione di livelli della competenza linguistica.

L'attenzione posta sull'insieme complesso di pratiche relative al *testing* linguistico è sollecitata dalla politica europea e da un sempre più esigente pubblico che aspira alla spendibilità delle proprie competenze in maniera trasparente e comprensibile su vasta scala. La panoramica offerta vuole far comprendere la rilevanza, ma anche la vastità del fenomeno presentato.

La ricerca ha voluto mostrare alcune teorie e metodologie di valutazione, per una adeguata conoscenza delle stesse, e alcune valide procedure statistiche che si prestano adeguatamente a un'indagine sul campo, per un'applicazione competente della stesse. I modelli illustrati sono apparsi degli imprescindibili strumenti di lavoro e di riferimento per l'intera analisi e si propongono come linee guida per tutti coloro che sono coinvolti nella valutazione linguistica.

Dal lavoro di ricerca emerge la complessità dell'intero sistema di elaborazione di una prova linguistica e degli strumenti di indagine che sono alla

base dell'analisi dei dati raccolti in seguito alla somministrazione di tale prova.

Il lavoro della prima parte vuole essere un contributo alla ricerca sul *testing* linguistico e il lavoro della seconda parte vuole offrire uno strumento pratico al personale impegnato nel campo dell'insegnamento e apprendimento delle lingue, e della valutazione e certificazione linguistica.

Conclusioni sul lavoro di analisi

L'analisi statistica dei risultati conseguiti dai circa 150 candidati sottoposti a esame ha condotto a risultati interessanti, capaci di fornire informazioni quantitative utili per poter intervenire qualitativamente sui contenuti dei test e dunque, di conseguenza, poter migliorare le procedure di valutazione e certificazione delle competenze linguistico-comunicative offerte dal CLA di Palermo.

Dall'indagine statistica si può rilevare indubbiamente un'alta percentuale di miglioramento della validità dei test, in linea con le formulazioni espresse in materia di valutazione.

Preme sottolineare però che i risultati della ricerca non hanno la pretesa di validità assoluta, sebbene mostrano una certa consistenza. I limiti della ricerca possono essere imputabili in primo luogo al campione ristretto, poiché un numero maggiore di candidati avrebbe permesso di giungere a conclusioni più fondate, e in secondo luogo al ridotto numero delle prove, giacché una più ampia gamma di *item* avrebbe consentito un'indagine maggiormente accurata dal punto di vista sia statistico che linguistico.

Si può affermare tuttavia, grazie all'evidenza dei dati, che i confini tra la

teoria e la pratica andrebbero considerevolmente sfumati, e la commistione tra i modelli teorici e le operazioni pratiche conduce a esiti di un certo interesse e utilità.

A questo proposito sembra opportuno riportare, concludendo, alcune parole note di Spolsky:

"[Language testing] (it) constantly forces practical and theoretical issues into fruitful tension. The needs of the tester regularly challenge the theorist, just as the findings of the theorist repeatedly tempt the tester."

(Spolsky, 1995:2-3)

Proposte di ricerca e applicazione della ricerca

La ricerca proposta potrebbe aprire possibili percorsi futuri per l'avvio di attività legate all'ambito della valutazione e certificazione linguistica. Le proposte di ricerca possono riguardare da una parte la formazione di una squadra di esperti per la supervisione di prove linguistiche e dall'altra la creazione di test linguistici per la certificazione di livelli successivi al B1, destinati agli insegnanti della scuola primaria.

La prima proposta riguarda la formazione di una équipe di esperti all'interno dell'università che possa offrire supporto e assistenza a coloro che sono coinvolti nella progettazione e nella valutazione di prove linguistiche. Un tale gruppo potrebbe divenire un punto di riferimento, in collaborazione con altri enti a livello nazionale, per sostenere scuole e centri linguistici universitari che si occupano di educazione e valutazione linguistica, e per favorire, a livello europeo, la standardizzazione e la comparazione delle certificazioni, lavorando a stretto contatto con organismi europei esistenti.

Il secondo progetto consiste nello sviluppo di una banca dati di prove

linguistiche di livello B2, sulla scia del piano formativo delle competenze linguistiche dei docenti della scuola primaria avviato dal MIUR. Ciò comporterebbe il conseguente avvio di corsi di formazione avanzati, destinati agli insegnanti di scuola primaria, mirati all'acquisizione di maggiori competenze linguistiche, e la valutazione delle stesse, con successivo conseguimento della relativa certificazione, la quale attesterebbe il raggiungimento di un livello superiore di abilità.

I possibili percorsi presentati mirano a perfezionare uno strumento che riveste un ruolo di fondamentale importanza nel campo dell'apprendimento e del *testing* linguistico.

APPENDICI

- ◆ Appendice A: Tabelle di riferimento estratte dal QCER
- ◆ Appendice B: Esempio di una prova PFLISP Sicilia
- ◆ Appendice C: Tabelle riassuntive dei dati raccolti

Tabelle di riferimento estratte dal QCER

Griglia di autovalutazione

		A1	A2	B1	B2	C1	C2
COMPRESIONE	Ascolto	Riesco a riconoscere parole che mi sono familiari ed espressioni molto semplici riferite a me stesso, alla mia famiglia e al mio ambiente, purché le persone parlino lentamente e chiaramente.	Riesco a capire espressioni e parole di uso molto frequente relative a ciò che mi riguarda direttamente (per esempio informazioni di base sulla mia persona e sulla mia famiglia, gli acquisti, l'ambiente circostante e il lavoro). Riesco ad afferrare l'essenziale di messaggi e annunci brevi, semplici e chiari.	Riesco a capire gli elementi principali in un discorso chiaro in lingua standard su argomenti familiari, che affronto frequentemente al lavoro, a scuola, nel tempo libero ecc. Riesco a capire l'essenziale di molte trasmissioni radiofoniche e televisive su argomenti di attualità o temi di mio interesse personale o professionale, purché il discorso sia relativamente lento e chiaro.	Riesco a capire discorsi di una certa lunghezza e conferenze e a seguire argomentazioni anche complesse purché il tema mi sia relativamente familiare. Riesco a capire la maggior parte dei notiziari e delle trasmissioni TV che riguardano fatti d'attualità e la maggior parte dei film in lingua standard.	Riesco a capire un discorso lungo anche se non è chiaramente strutturato e le relazioni non vengono segnalate, ma rimangono implicite. Riesco a capire senza troppo sforzo le trasmissioni televisive e i film.	Non ho nessuna difficoltà a capire qualsiasi lingua parlata, sia dal vivo sia trasmessa, anche se il discorso è tenuto in modo veloce da un madrelingua, purché abbia il tempo di abituarci all'accento.
	Lettura	Riesco a capire i nomi e le persone che mi sono familiari e frasi molto semplici, per esempio quelle di annunci, cartelloni, cataloghi.	Riesco a leggere testi molto brevi e semplici e a trovare informazioni specifiche e prevedibili in materiale di uso quotidiano, quali pubblicità, programmi, menù e orari. Riesco a capire lettere personali semplici e brevi.	Riesco a capire testi scritti di uso corrente legati alla sfera quotidiana o al lavoro. Riesco a capire la descrizione di avvenimenti, di sentimenti e di desideri contenuta in lettere personali.	Riesco a leggere articoli e relazioni su questioni d'attualità in cui l'autore prende posizione ed esprime un punto di vista determinato. Riesco a comprendere un testo narrativo contemporaneo.	Riesco a capire testi letterari e informativi lunghi e complessi e so apprezzare le differenze di stile. Riesco a capire articoli specialistici e istruzioni tecniche piuttosto lunghe, anche quando non appartengono al mio settore.	Riesco a capire con facilità praticamente tutte le forme di lingua scritta inclusi i testi teorici, strutturalmente o linguisticamente complessi, quali manuali, articoli specialistici e opere letterarie.
PARLATO	Interazione	Riesco a interagire in modo semplice se l'interlocutore è disposto a ripetere o a riformulare più lentamente certe cose e mi aiuta a formulare ciò che cerco di dire. Riesco a porre e a rispondere a domande semplici su argomenti molto familiari o che riguardano bisogni immediati.	Riesco a comunicare affrontando compiti semplici e di routine che richiedano solo uno scambio semplice e diretto di informazioni su argomenti e attività consuete. Riesco a partecipare a brevi conversazioni, anche se di solito non capisco abbastanza per riuscire a sostenere la conversazione.	Riesco ad affrontare molte delle situazioni che si possono presentare viaggiando in una zona dove si parla la lingua. Riesco a partecipare, senza essermi preparato, a conversazioni su argomenti familiari, di interesse personale o riguardanti la vita quotidiana (per esempio la famiglia, gli hobby, il lavoro, i viaggi e i fatti di attualità).	Riesco a comunicare con un grado di spontaneità e scioltezza sufficiente per interagire in modo normale con parlanti nativi. Riesco a partecipare attivamente a una discussione in contesti familiari, esponendo e sostenendo le mie opinioni.	Riesco ad esprimermi in modo sciolto e spontaneo senza dover cercare troppo le parole. Riesco ad usare la lingua in modo flessibile ed efficace nelle relazioni sociali e professionali. Riesco a formulare idee e opinioni in modo preciso e a collegare abilmente i miei interventi con quelli di altri interlocutori.	Riesco a partecipare senza sforzi a qualsiasi conversazione e discussione ed ho familiarità con le espressioni idiomatiche e colloquiali. Riesco ad esprimermi con scioltezza e a rendere con precisione sottili sfumature di significato. In caso di difficoltà, riesco a ritornare sul discorso e a riformularlo in modo così scorrevole che difficilmente qualcuno se ne accorge.

S C R I T T O	Produzione orale	Riesco a usare espressioni e frasi semplici per descrivere il luogo dove abito e la gente che conosco.	Riesco ad usare una serie di espressioni e frasi per descrivere con parole semplici la mia famiglia ed altre persone, le mie condizioni di vita, la carriera scolastica e il mio lavoro attuale o il più recente.	Riesco a descrivere, collegando semplici espressioni, esperienze ed avvenimenti, i miei sogni, le mie speranze e le mie ambizioni. Riesco a motivare e spiegare brevemente opinioni e progetti. Riesco a narrare una storia e la trama di un libro o di un film e a descrivere le mie impressioni.	Riesco a esprimermi in modo chiaro e articolato su una vasta gamma di argomenti che mi interessano. Riesco a esprimere un' opinione su un argomento d'attualità, indicando vantaggi e svantaggi delle diverse opzioni.	Riesco a presentare descrizioni chiare e articolate su argomenti complessi, integrandovi temi secondari, sviluppando punti specifici e concludendo il tutto in modo appropriato.	Riesco a presentare descrizioni o argomentazioni chiare e scorrevoli, in uno stile adeguato al contesto e con una struttura logica efficace, che possa aiutare il destinatario a identificare i punti salienti da rammentare.
	Produzione scritta	Riesco a scrivere una breve e semplice cartolina , ad esempio per mandare i saluti delle vacanze. Riesco a compilare moduli con dati personali scrivendo per esempio il mio nome, la nazionalità e l'indirizzo sulla scheda di registrazione di un albergo.	Riesco a prendere semplici appunti e a scrivere brevi messaggi su argomenti riguardanti bisogni immediati. Riesco a scrivere una lettera personale molto semplice, per esempio per ringraziare qualcuno.	Riesco a scrivere testi semplici e coerenti su argomenti a me noti o di mio interesse. Riesco a scrivere lettere personali esponendo esperienze e impressioni.	Riesco a scrivere testi chiari e articolati su un'ampia gamma di argomenti che mi interessano. Riesco a scrivere saggi e relazioni, fornendo informazioni e ragioni a favore o contro una determinata opinione. Riesco a scrivere lettere mettendo in evidenza il significato che attribuisco personalmente agli avvenimenti e alle esperienze.	Riesco a scrivere testi chiari e ben strutturati sviluppando analiticamente il mio punto di vista. Riesco a scrivere lettere, saggi e relazioni esponendo argomenti complessi, evidenziando i punti che ritengo salienti. Riesco a scegliere lo stile adatto ai lettori ai quali intendo rivolgermi.	Riesco a scrivere testi chiari, scorrevoli e stilisticamente appropriati. Riesco a scrivere lettere, relazioni e articoli complessi, supportando il contenuto con una struttura logica efficace che aiuti il destinatario a identificare i punti salienti da rammentare. Riesco a scrivere riassunti e recensioni di opere letterarie e di testi specialisti.

Tabella n. 1 - Livelli comuni di riferimento: griglia di autovalutazione (QCER:34-35)

Domini

Ambito	Luoghi	Istituzioni	Persone	Oggetti	Avvenimenti	Azioni	Testi
Privato	Casa (appartamento, locali, giardino) propria della famiglia di amici di estranei Spazio per sé in una pensione, in un hotel Campagna, mare	La famiglia Reti sociali	Genitori Nonni e nonne Figli e nipoti Fratelli e sorelle Zii e zie Cugini e cugine, affini, Coniugi, persone con cui si è in intimità Amici, conoscenti	Mobili e arredo Abbigliamento Apparecchi domestici Giocattoli, attrezzi, igiene personale Oggetti d'arte, libri Animali (domestici) Alberi, piante, prati, stagni Beni domestici Bagagli Attrezzature per il tempo libero e lo sport	Feste in famiglia Incontri Incidenti, infortuni Fenomeni naturali Feste, visite Passeggiate a piedi, in bicicletta, in moto Vacanze, escursioni Eventi sportivi	Routine quotidiane: vestirsi, svestirsi, cucinare, mangiare, lavarsi Bricolage, giardinaggio Lettura, radio e TV Divertimenti Hobby Giochi e sport	Telex Garanzie Ricette Manuali scolastici Romanzi, riviste Giornali Materiale pubblicitario Opuscoli Lettere personali Testi orali diffusi via radio e registrati
Pubblico	Luoghi pubblici: strada, piazza, parco Mezzi di trasporto Negozi, (super)mercati Ospedali, ambulatori cliniche Stadi, campi sportivi, palestre Teatri, cinema spettacoli Ristoranti, bar, alberghi, Luoghi di culto	Uffici statali Organismi politici Organismi giudiziari Servizio sanitario Associazioni di volontariato Gruppi Partiti politici Istituzioni religiose	Semplici cittadini Rappresentanti ufficiali Commessi Polizia, esercito, personale di sicurezza Autisti, controllori Passeggeri Giocatori, fans, spettatori Attori, pubblico, Camerieri, personale dei bar Portieri Preti, fedeli	Soldi, portamonete, portafogli Documenti Merci Armi Zaini, valige, borse Palloni Programmi Pasti, bevande, merende Passaporti, patenti	Incidenti Infortuni, malattie Incontri pubblici Processi, udienze in tribunale Giornate benefiche, Multe, arresti Partite, gare Spettacoli Matrimoni, funerali	Acquistare i servizi pubblici e utilizzarli Usare il servizio sanitario Viaggi in auto, treno, nave, aereo Divertimenti e attività del tempo libero Funzioni religiose	Annunci e avvisi Etichette, confezioni Volantini, graffiti Biglietti, orari Cartelli, regolamenti Programmi Contratti Menu Testi sacri, prediche, preghiere
Educativo	Scuole: atrio, aule, cortile, campi sportivi, corridoi Collegi Università Associazioni culturali Sale per le conferenze Aule per i seminari Associazioni studentesche Residenze universitarie Laboratori Mense	Scuola Collegio Università Associazioni culturali Associazioni professionali Organismi di formazione continua	Insegnanti Bidelli Assistenti Genitori Compagni di classe Professori universitari, lettori Studenti universitari Bibliotecari e personale di laboratorio Personale della mensa Personale delle pulizie Portieri, segretari	Materiale per scrivere Abbigliamento e attrezzature sportive Cibo Apparecchi audiovisivi Lavagna e gesso Computer Cartelle e zaini	Inizio d'anno Ingresso in classe Fine anno Visite e scambi Pomeriggi e serate con i genitori Giornate sportive, partite Problemi disciplinari	Lezioni Giochi Ricreazione Gruppi e associazioni Compiti in classe Lavori in laboratorio Studio in biblioteca Seminari e lavori assistiti Compiti a casa Dibattiti e discussioni	Testi autentici (vedi sopra) Libri di testo, di lettura Libri di consultazione Testi alla lavagna (luminosa) Testi su schermo (computer e TV) Esercizi Articoli di giornale Sommari Dizionari

Tabella n. 2 - Domini o ambiti d'interesse (QCER:60-61)

Tipologie di valutazione

valutazione del profitto	consiste nell'accertare se sono stati raggiunti obiettivi specifici in base a ciò che è stato insegnato.
valutazione della competenza	consiste nell'appurare ciò che un individuo sa fare/conosce con riferimento al mondo reale.
valutazione riferita alla norma	consiste nel valutare l'esito della prova di un individuo classificandola in una graduatoria rispetto a quelle dei suoi pari.
valutazione riferita al criterio	consiste nel valutare esclusivamente la capacità di un individuo in un ambito specifico, senza alcun riferimento ai suoi pari.
valutazione in base a criteri di padronanza	consiste nel discriminare, mediante un preciso <i>standard</i> di competenza minima o di soglia di accettabilità, tra apprendenti che hanno padronanza e apprendenti che non ne hanno.
valutazione in base a criteri di progressività	consiste nel valutare la capacità di un individuo in base a un continuum che comprende tutti i livelli significativi per quella determinata capacità.
valutazione continua	consiste nella valutazione attuata dall'insegnante e/o dallo studente sul rendimento e sui progetti realizzati durante tutto il corso.
valutazione puntuale	consiste nella valutazione attuata in un giorno preciso, solitamente alla fine o prima dell'inizio di un corso, tramite un esame per l'assegnazione di un voto e/o per la determinazione di una decisione.
valutazione formativa	consiste in un processo continuo che permette di raccogliere informazioni sull'apprendimento, sui punti di forza e di debolezza da considerare nella programmazione e da comunicare all'apprendente.
Valutazione sommativa	consiste nel riassumere in un voto i risultati raggiunti alla fine di un corso.
valutazione diretta	consiste nel valutare ciò che il candidato sta facendo (ad es. valutazione del parlato, scritto e ascolto tramite prove in classe).
valutazione indiretta	consiste nel valutare le abilità potenziali ricorrendo a una prova (ad es. valutazione della lettura tramite la prova di comprensione).
valutazione delle prestazioni	consiste nel valutare un campione di lingua parlata o scritta fornito dall'apprendente durante una prova diretta.
valutazione delle conoscenze	consiste nel valutare tramite deduzione l'ampiezza e la padronanza delle conoscenze linguistiche di un individuo mediante prove di vario genere.
valutazione soggettiva	consiste nel manifestare un giudizio sulla qualità di una prestazione.
valutazione oggettiva	consiste nell'espressione di un giudizio privo di soggettività (di solito si tratta di valutare prove indirette in cui ogni domanda ha un'unica risposta corretta)
classificare in base a una scala	consiste nel classificare un individuo che si trova a un livello specifico entro una determinata fascia su una scala numerica a più livelli.
classificare in base a una lista di controllo	consiste nel valutare una persona riferendosi a una lista articolata di punti considerati rilevanti per un determinato livello.
giudizio in base all'impressione soggettiva	consiste in un giudizio del tutto soggettivo formulato senza alcun riferimento a criteri specifici
giudizio in base a criteri	consiste in un giudizio formulato attraverso il riferimento a criteri specifici che limitano la soggettività.
valutazione olistica	consiste nell'espressione di un giudizio globale sintetico.
valutazione analitica	consiste nella valutazione di alcuni aspetti attraverso un'osservazione minuziosa della prestazione.
valutazione di categorie	consiste nel valutare una singola prova di verifica sulla base di categorie definite in una griglia.
valutazione di serie	consiste nel valutare un certo numero di prove distinte con un solo voto sulla base di una scala olistica.
eterovalutazione	consiste nella valutazione attuata dall'insegnante o dall'esaminatore.
autovalutazione	consiste nella valutazione espressa dall'individuo stesso sulla propria competenza.

Tabella n. 3 – Tipologie di valutazione (QCER:224-236)

Esempio di una prova PFLISP Sicilia

Centro Linguistico di Ateneo
Università degli Studi di Palermo
ESAME PFLISP SICILIA

Prova di Listening

Tempo: 30 minuti

Punteggio max: 20

Modalità di somministrazione: supporto informatico

Prova di Reading

Tempo: 30 minuti

Punteggio max: 25

Modalità di somministrazione: supporto informatico

Prova di Writing

Tempo: 45 minuti

Punteggio max: 25

Modalità di somministrazione: supporto cartaceo

Prova di Speaking

Tempo: 10 minuti

Punteggio max: 30

Modalità di somministrazione: colloquio

Istruzioni per il candidato:

Per le prove su supporto informatico accedere alla piattaforma Anacleto inserendo il proprio username e la propria password.

Per la prova su supporto cartaceo scrivere il proprio nome e cognome e la data odierna nella parte superiore del foglio.

Informazioni per il candidato:

Ogni risposta omessa o errata vale 0.

Listening

La prova è costituita da 2 parti. Ogni domanda vale due punti. Ogni dialogo può essere ascoltata tre volte.

Reading

La prova è costituita da 3 parti. Ogni domanda vale un punto.

Writing

La prova è costituita da 2 parti. La parte prima vale 15 punti. La parte seconda vale 10 punti.

PROVA DI LISTENING

PRIMA PARTE

There are five questions in this part.
You will hear a short recording for each question.
Choose the correct answer.

Jane: I'm hungry. What about a snack.
Tom: That's a good idea! Let's go to the cafeteria.

1. Jane and Tom are:
- a) cold.
 - b) hungry.
 - c) in a hurry.

Jane: Oh, look! There's Care and Paula.
Tom : Who are they?
Jane : They're my friends

2. Jane and Tom meet:
- a) two friends.
 - b) two men.
 - c) two neighbours.

Claire: Here I am Tom. Sorry, but my car broke down. I had to take a taxi.
Tom: Never mind. I was looking at the windows.

3. Claire is:
- a) punctual.
 - b) late.
 - c) early.

Tom: Sandra look! That's a nice dress, there. The black one with a white collar.
Sandra: Yes. Beautiful. But I am looking for a shirt , actually.
Tom: Well, let's go inside and have a look.

4. Sandra is looking for:
- a) a sweater.
 - b) a shirt.
 - c) a skirt.

Tom: What are you doing on Sunday night? We can go to the cinema.
Ellis: Sorry, I can't. I've got a maths test on Monday morning, so I have to study hard.

5. Ellis has:
- a) a lot to do on Sunday.
 - b) nothing to do on Sunday.
 - c) a lot of spare time.

SECONDA PARTE

Listen to the dialogue between a man and a woman talking about a job interview and decide if each statement is true or false.

Woman: Are you going for that job interview tomorrow?
 Man: Yes, I am. I have to go at ten o' clock in the morning. I don't know how to get there, though.
 Woman: I can take you there in my car, if you like.
 Man: Thanks a lot. That's a great help. I must say I'm not happy about this interview.
 Woman: Why not, you have got the right qualifications. And I know they need engineers
 Man: I know, but I haven't had much experience.
 Woman: I don't think that is important. I'm sure you'll get the job
 Man: I don't know. I don't feel very confident about it.
 Woman: Do you want the job?
 Man: Of course I do. I really need it.
 Woman: Does it involve a lot of travelling?
 Man: About three months a year. And that's another thing. I'm not very good at languages.
 Woman: But your German is excellent. I was very impressed with the way you spoke to those Germans we met on holiday last year.
 Man: But I don't speak it well enough for doing business, and my French is terrible.
 Woman: Well, I don't agree. I think you speak German and French well.
 Man: Thank you for trying to make me feel better but I still don't think I'll get the job.
 Woman: Well. We'll just have to wait and see, won't we?

- 1 The man is looking forward to the interview.
- 2 The man must travel if he gets the job.
- 3 The woman was impressed by the man's German.
- 4 The man really wants the job.
- 5 The man feels confident about his ability to speak foreign languages.

T	F
T	F
T	F
T	F
T	F

PROVA DI READING**PRIMA PARTE**

Read the story about a person who went on a sailing trip and then answer the questions by choosing the best answer a, b or c.

Sailing to Morocco

I love all water sports. I do diving in the summer and rowing and canoeing in the winter but my favourite is sailing on my father's boat. I've travelled all over the world with my family. To tell the truth I've forgotten the names of most places we've visited. Some places stick in my mind though – like our first trip to North Africa.

Sailing from Spain to Morocco was a bit boring at first. However, the best part of the whole journey was during our windy day when our boat was breaking through the huge waves. We were going quite fast. It was so exciting and such a thrill to be experiencing nature in this way.

My little brother on the other hand, was terrified and started to cry. All of a sudden, I spotted five or six dolphins diving in and out the waves, not far from the boat. When my little brother saw them his face lit up – he was delighted and he forgot about being scared.

Later in the afternoon, we reached Casablanca. I was so relieved and happy to be on land, so were the rest of the family. It was such a busy place. I wanted to go for a walk, but I was so tired that I fell asleep instead.

1. The narrator prefers:

- a) diving and rowing.
- b) rowing and canoeing.
- c) going on sailing boats.

2. How many of the places he has visited can he remember?

- a) All of them.
- b) Most of them.
- c) Some of them.

3. Sailing from Spain to Morocco overall was a thrill because:

- a) the boat went slowly.
- b) the wind was strong.
- c) he helped his father.

4. When the narrator's little brother saw the dolphins he was:

- a) terrified.
- b) calm.
- c) worried.

5. At the end of the journey:
- a) the narrator fell asleep.
 - b) they were all glad to be ashore.
 - c) the narrator went for a walk.

SECONDA PARTE

Read the passage and then answer the questions by choosing the best answer a, b or c.

Thanksgiving Day

Thanksgiving Day is a truly American holiday. It was first celebrated in 1621 by Pilgrim Fathers, some English colonists who set up the colony of Massachusetts. They came to America on the famous ship, the Mayflower, because they wanted to have religious freedom.

The first winter in the new land was very difficult and a lot of the colonists died. But some friendly Native Americans helped the Pilgrim Fathers. In fact, they taught the English colonists how to plant corn, and how to hunt animals for food. At the end of the summer, they produced a lot of food and the Governor of the colony called a day of thanksgiving in November to thank God for the good harvest. He invited the Native Americans to help them celebrate.

The colonists prepared a big meal. They ate roast turkey and a lot of other things like pumpkins, corn and fruit. The feast lasted three days.

Today, Americans celebrate Thanksgiving on the fourth Thursday of November and they eat almost the same things as the Pilgrim Fathers ate on that first Thanksgiving Day in 1621: roast turkey, corn dishes and pumpkin pie.

1. Thanksgiving Day is:
- a) only an English holiday.
 - b) an English and American holiday.
 - c) only an American holiday.
2. The first Thanksgiving Day was about:
- a) 400 years ago.
 - b) 200 years ago.
 - c) 500 years ago.
3. Thanksgiving Day celebrates:
- a) the production of a lot of food.
 - b) the discovery of corn.
 - c) the arrival of English colonists in America.
4. The Pilgrim Fathers invited Native Americans to celebrate with them because:
- a) the colonists wanted to help the Natives.
 - b) the colonists wanted to show their gratitude to Natives.

- c) the Natives were not friendly.

5. Americans today:

- a) celebrate Thanksgiving in November.
- b) don't celebrate Thanksgiving.
- c) celebrate Thanksgiving but they don't eat any food.

TERZA PARTE

Read the following article and decide if the sentences are true or false.

Doctors Without Frontiers

Kelli Miller was a very successful doctor in England. For the last six months she has been living in Ethiopia, working for an organisation called Doctors Without Frontiers that provides emergency health care in African countries.

Kelli was bored with treating patients with routine medical problems. She feels that her job in Ethiopia is much more important. There aren't enough hospitals or medical equipment so the health care standards are very low. Kelli has been working very hard to improve conditions in Ethiopia and at the moment she is training local people as health care workers.

Kelli says that helping people who really need it has made her feel more useful than ever before. She will be staying in Ethiopia for a very long time.

- 1 Kelli Miller has always worked in Ethiopia.
- 2 Kelli found her job in England very interesting.
- 3 Health care conditions in Ethiopia are very bad.
- 4 Kelli will leave Ethiopia as soon as possible.
- 5 The article is informing about English doctors living in Africa.

T	F
T	F
T	F
T	F
T	F

QUARTA PARTE

Read the text below and choose the correct item for each space.

The explorer David Livingstone was (1).....in Scotland in 1813. At the (2)..... of ten, he was sent (3)..... work in the local cotton factory. His working day began at six o'clock in the morning and didn't end (4).....eight o'clock in the evening.

(5)..... Livingstone was a young man, he became interested in missionary work. On November 20, 1840, he sailed (6).....Africa to work in a settlement. Once there, he became very popular (7).....the natives, treating them (8).....equals. With their help, he travelled across the great continent of Africa. On his travels, he made many discoveries. On one exciting journey, he (9).....across the great Victoria Waterfall. (10).....the end of his life, Livingstone was overworked and in poor health. He died in April 1873, in a small village in Africa.

1	A lived	B born	C bring up
2	A age	B period	C month
3	A for	B to	C from
4	A before	B till	C at
5	A When	B As	C Because
6	A in	B to	C at
7	A between	B around	C with
8	A as	B more	C very
9	A sailed	B drove	C walked
10	A At	B In	C After

PROVA DI WRITING

PRIMA PARTE

Write an article about a famous beach campsite reporting about:

- The name and location of the place, why it is special.
- Description of the place and activities you can do there.
- Recommendation and best time to go.

Use about (100 words)

SECONDA PARTE

Here are some sentences.

Complete the second sentence so that it means the same as the first .

0. E.g. You can buy your tickets in advance.

Tickets can be bought in advance.

1. When was the last time you went to the cinema?
When did.....
2. It is not necessary to pay the entire price.
You.....
3. Have you had any news from John?
Have you heard.....
4. Shall we play tennis?
How about.....
5. Have you got a scooter?
Do.....
6. I advise you to lose weight.
You.....
7. "When are you starting working as a bank manager?" Paul asked Bill.
Paul asked Bill.....
8. As he is too young, he can't enter the competition.
He can't enter the competition.....
9. Scientists often make new discoveries all the time.
New discoveries.....
10. Each house has four rooms.
.....**four rooms in each house.**

Nome e Cognome del candidato _____ Data _____

Tabelle riassuntive dei dati raccolti

Esiti prova di Listening e Reading n. 1 Gruppo A (35 candidati)

Candidato	Prova di ascolto Punteggio su base 20	Prova di lettura Punteggio su base 25
Candidato n. 1	18	24
Candidato n. 2	18	19
Candidato n. 3	18	22
Candidato n. 4	18	21
Candidato n. 5	18	21
Candidato n. 6	16	20
Candidato n. 7	12	17
Candidato n. 8	16	18
Candidato n. 9	16	17
Candidato n. 10	16	21
Candidato n. 11	12	19
Candidato n. 12	18	22
Candidato n. 13	18	22
Candidato n. 14	18	21
Candidato n. 15	18	21
Candidato n. 16	16	18
Candidato n. 17	16	22
Candidato n. 18	18	23
Candidato n. 19	14	17
Candidato n. 20	12	20
Candidato n. 21	14	13
Candidato n. 22	14	17
Candidato n. 23	16	13
Candidato n. 24	16	15
Candidato n. 25	18	22
Candidato n. 26	16	23
Candidato n. 27	16	23
Candidato n. 28	16	22
Candidato n. 29	16	20
Candidato n. 30	16	20
Candidato n. 31	16	20
Candidato n. 32	16	21
Candidato n. 33	16	20
Candidato n. 34	18	22
Candidato n. 35	12	18

Listening

Gruppo A *higher* (10 candidati):

Candidato	Prova di ascolto (punteggio su base 20)
Candidato n. 1	18
Candidato n. 2	18
Candidato n. 3	18
Candidato n. 4	18
Candidato n. 5	18
Candidato n. 12	18
Candidato n. 13	18
Candidato n. 14	18
Candidato n. 15	18
Candidato n. 18	18

Gruppo A *lower* (10 candidati):

Candidato	Prova di ascolto (punteggio su base 20)
Candidato n. 31	16
Candidato n. 32	16
Candidato n. 33	16
Candidato n. 19	14
Candidato n. 21	14
Candidato n. 22	14
Candidato n. 7	12
Candidato n. 11	12
Candidato n. 20	12
Candidato n. 35	12

Reading

Gruppo A *higher* (10 candidati):

Candidato	Prova di lettura (punteggio su base 25)
Candidato n. 1	24
Candidato n. 18	23
Candidato n. 26	23
Candidato n. 27	23
Candidato n. 3	22
Candidato n. 12	22
Candidato n. 13	22
Candidato n. 17	22
Candidato n. 25	22
Candidato n. 28	22

Gruppo A *lower* (10 candidati):

Candidato	Prova di lettura (punteggio su base 25)
Candidato n. 8	18
Candidato n. 16	18
Candidato n. 35	18
Candidato n. 7	17
Candidato n. 9	17
Candidato n. 19	17
Candidato n. 22	17
Candidato n. 24	15
Candidato n. 21	13
Candidato n. 23	13

Listening

Deviazione standard Gruppo A (35 candidati)

Punteggi su base 20	f	fx	x-m	(x-M) ²	f (x-M) ²
18	12	216	1,95	3,80	45,63
16	16	256	-0,05	0,00	0,04
14	3	42	-2,05	4,20	12,61
12	4	48	-4,05	16,40	65,61
	$\Sigma = 35$	M = 16,05			$\Sigma = 123,89$

Reading

Deviazione standard Gruppo A (35 candidati)

Punteggi su base 25	f	fx	x-m	(x-M) ²	f (x-M) ²
24	1	24	4,18	17,47	17,47
23	3	69	3,18	10,11	30,34
22	7	154	2,18	4,75	33,27
21	6	126	1,18	1,39	8,35
20	6	120	0,18	0,03	0,19
19	2	38	-0,82	0,67	1,34
18	3	54	-1,82	3,31	9,94
17	4	68	-2,82	7,95	31,81
15	1	15	-4,82	23,23	23,23
13	2	26	-6,82	46,51	93,02
	$\Sigma = 35$	M = 19,82			$\Sigma = 248,97$

Listening
Punteggi Z e T Gruppo A (35 candidati)

Graduatoria	Candidato	Punteggio su base 20	Punteggio Z	Punteggio T
6,5	Candidato n. 1	18	1,04	60,36
6,5	Candidato n. 2	18	1,04	60,36
6,5	Candidato n. 3	18	1,04	60,36
6,5	Candidato n. 4	18	1,04	60,36
6,5	Candidato n. 5	18	1,04	60,36
6,5	Candidato n. 12	18	1,04	60,36
6,5	Candidato n. 13	18	1,04	60,36
6,5	Candidato n. 14	18	1,04	60,36
6,5	Candidato n. 15	18	1,04	60,36
6,5	Candidato n. 18	18	1,04	60,36
6,5	Candidato n. 25	18	1,04	60,36
6,5	Candidato n. 34	18	1,04	60,36
20,5	Candidato n. 6	16	-0,03	49,73
20,5	Candidato n. 8	16	-0,03	49,73
20,5	Candidato n. 9	16	-0,03	49,73
20,5	Candidato n. 10	16	-0,03	49,73
20,5	Candidato n. 16	16	-0,03	49,73
20,5	Candidato n. 17	16	-0,03	49,73
20,5	Candidato n. 23	16	-0,03	49,73
20,5	Candidato n. 24	16	-0,03	49,73
20,5	Candidato n. 26	16	-0,03	49,73
20,5	Candidato n. 27	16	-0,03	49,73
20,5	Candidato n. 28	16	-0,03	49,73
20,5	Candidato n. 29	16	-0,03	49,73
20,5	Candidato n. 30	16	-0,03	49,73
20,5	Candidato n. 31	16	-0,03	49,73
20,5	Candidato n. 32	16	-0,03	49,73
20,5	Candidato n. 33	16	-0,03	49,73
30	Candidato n. 19	14	-1,09	39,10
30	Candidato n. 21	14	-1,09	39,10
30	Candidato n. 22	14	-1,09	39,10
33,5	Candidato n. 7	12	-2,15	28,47
33,5	Candidato n. 11	12	-2,15	28,47
33,5	Candidato n. 20	12	-2,15	28,47
33,5	Candidato n. 35	12	-2,15	28,47

Reading
Punteggi Z e T Gruppo A (35 candidati)

Graduatoria	Candidato	Punteggio su base 25	Punteggi Z	Punteggi T
1	Candidato n. 1	24	1,57	65,67
3	Candidato n. 18	23	1,19	61,92
3	Candidato n. 26	23	1,19	61,92
3	Candidato n. 27	23	1,19	61,92
8	Candidato n. 3	22	0,82	58,17
8	Candidato n. 12	22	0,82	58,17
8	Candidato n. 13	22	0,82	58,17
8	Candidato n. 17	22	0,82	58,17
8	Candidato n. 25	22	0,82	58,17
8	Candidato n. 28	22	0,82	58,17
8	Candidato n. 34	22	0,82	58,17
14,5	Candidato n. 4	21	0,44	54,42
14,5	Candidato n. 5	21	0,44	54,42
14,5	Candidato n. 10	21	0,44	54,42
14,5	Candidato n. 14	21	0,44	54,42
14,5	Candidato n. 15	21	0,44	54,42
14,5	Candidato n. 32	21	0,44	54,42
20,5	Candidato n. 6	20	0,07	50,67
20,5	Candidato n. 20	20	0,07	50,67
20,5	Candidato n. 29	20	0,07	50,67
20,5	Candidato n. 30	20	0,07	50,67
20,5	Candidato n. 31	20	0,07	50,67
20,5	Candidato n. 33	20	0,07	50,67
24,5	Candidato n. 2	19	-0,31	46,93
24,5	Candidato n. 11	19	-0,31	46,93
27	Candidato n. 8	18	-0,68	43,18
27	Candidato n. 16	18	-0,68	43,18
27	Candidato n. 35	18	-0,68	43,18
30,5	Candidato n. 7	17	-1,06	39,43
30,5	Candidato n. 9	17	-1,06	39,43
30,5	Candidato n. 19	17	-1,06	39,43
30,5	Candidato n. 22	17	-1,06	39,43
33	Candidato n. 24	15	-1,81	31,93
34,5	Candidato n. 21	13	-2,56	24,43
34,5	Candidato n. 23	13	-2,56	24,43

Listening e Reading
Differenze di posizione Gruppo A (35 candidati)

Candidato	Ascolto		Lettura		D	D ²
	Graduatoria	Punteggio T	Graduatoria	Punteggi T		
Candidato n. 01	6,5	60,36	1	65,67	5,5	30,25
Candidato n. 02	6,5	60,36	24,5	46,93	-18	324
Candidato n. 03	6,5	60,36	8	58,17	-1,5	2,25
Candidato n. 04	6,5	60,36	14,5	54,42	-8	64
Candidato n. 05	6,5	60,36	14,5	54,42	-8	64
Candidato n. 06	20,50	49,73	20,5	50,67	0	0
Candidato n. 07	33,5	28,47	30,5	39,43	3	9
Candidato n. 08	20,5	49,73	27	43,18	-6,5	42,25
Candidato n. 09	20,5	49,73	30,5	39,43	-10	100
Candidato n. 10	20,5	49,73	14,5	54,42	6	36
Candidato n. 11	33,5	28,47	24,5	46,93	9	81
Candidato n. 12	6,5	60,36	8	58,17	-1,5	2,25
Candidato n. 13	6,5	60,36	8	58,17	-1,5	2,25
Candidato n. 14	6,5	60,36	14,5	54,42	-8	64
Candidato n. 15	6,5	60,36	14,5	54,42	-8	64
Candidato n. 16	20,5	49,73	27	43,18	-6,5	42,25
Candidato n. 17	20,5	49,73	8	58,17	12,5	156,25
Candidato n. 18	6,5	60,36	3	61,92	3,5	12,25
Candidato n. 19	30	39,1	30,5	39,43	-0,5	0,25
Candidato n. 20	33,5	28,47	20,5	50,67	13	169
Candidato n. 21	30	39,1	34,5	24,43	-4,5	20,25
Candidato n. 22	30	39,1	30,5	39,43	-0,5	0,25
Candidato n. 23	20,5	49,73	34,5	24,43	-14	196
Candidato n. 24	20,5	49,73	33	31,93	-12,5	156,25
Candidato n. 25	6,5	60,36	8	58,17	-1,5	2,25
Candidato n. 26	20,5	49,73	3	61,92	17,5	306,25
Candidato n. 27	20,5	49,73	3	61,92	17,5	306,25
Candidato n. 28	20,5	49,73	8	58,17	12,5	156,25
Candidato n. 29	20,5	49,73	20,5	50,67	0	0
Candidato n. 30	20,5	49,73	20,5	50,67	0	0
Candidato n. 31	20,5	49,73	20,5	50,67	0	0
Candidato n. 32	20,5	49,73	14,5	54,42	6	36
Candidato n. 33	20,5	49,73	20,5	50,67	0	0
Candidato n. 34	6,5	60,36	8	58,17	-1,5	2,25
Candidato n. 35	33,5	28,47	27	43,18	6,5	42,25
						Σ = 2489,5

Esiti prova di Listening e Reading n. 2

Gruppo B (34 candidati)

Candidato	Prova di ascolto Punteggio su base 20	Prova di lettura Punteggio su base 25
Candidato n. 36	16	19
Candidato n. 37	14	13
Candidato n. 38	20	14
Candidato n. 39	16	18
Candidato n. 40	20	14
Candidato n. 41	12	12
Candidato n. 42	18	11
Candidato n. 43	16	15
Candidato n. 44	12	8
Candidato n. 45	20	13
Candidato n. 46	20	15
Candidato n. 47	14	9
Candidato n. 48	18	8
Candidato n. 49	8	5
Candidato n. 50	20	15
Candidato n. 51	20	13
Candidato n. 52	16	16
Candidato n. 53	10	9
Candidato n. 54	20	14
Candidato n. 55	14	10
Candidato n. 56	16	9
Candidato n. 57	14	6
Candidato n. 58	16	15
Candidato n. 59	14	18
Candidato n. 60	14	18
Candidato n. 61	16	20
Candidato n. 62	10	19
Candidato n. 63	18	12
Candidato n. 64	12	19
Candidato n. 65	16	12
Candidato n. 66	12	15
Candidato n. 67	16	14
Candidato n. 68	16	14
Candidato n. 69	16	19

Listening

Gruppo B *higher* (9 candidati):

Candidato	Prova di ascolto (punteggio su base 20)
Candidato n. 38	20
Candidato n. 40	20
Candidato n. 45	20
Candidato n. 46	20
Candidato n. 50	20
Candidato n. 51	20
Candidato n. 54	20
Candidato n. 42	18
Candidato n. 48	18

Gruppo B *lower* (9 candidati):

Candidato	Prova di ascolto (punteggio su base 20)
Candidato n. 55	14
Candidato n. 57	14
Candidato n. 59	14
Candidato n. 60	14
Candidato n. 41	12
Candidato n. 44	12
Candidato n. 64	12
Candidato n. 66	12
Candidato n. 53	10

Reading

Gruppo B *higher* (9 candidati):

Candidato	Prova di lettura (punteggio su base 25)
Candidato n. 61	20
Candidato n. 36	19
Candidato n. 62	19
Candidato n. 64	19
Candidato n. 69	19
Candidato n. 39	18
Candidato n. 50	18
Candidato n. 60	18
Candidato n. 52	16

Gruppo B *lower* (9 candidati):

Candidato	Prova di lettura (punteggio su base 25)
Candidato n. 42	11
Candidato n. 55	10
Candidato n. 47	9
Candidato n. 53	9
Candidato n. 56	9
Candidato n. 44	8
Candidato n. 48	8
Candidato n. 57	6
Candidato n. 49	5

Listening

Deviazione standard Gruppo B (34 candidati)

Punteggi su base 20	f	fx	x-m	(x-M) ²	f (x-M) ²
20	7	140	4,42	19,54	136,75
18	3	54	2,42	5,86	17,57
16	11	176	0,42	0,18	1,94
14	6	84	-1,58	2,50	14,98
12	4	48	-3,58	12,82	51,27
10	2	20	-5,58	31,14	62,27
8	1	8	-7,58	57,46	57,46
	$\Sigma = 34$	M = 15,58			$\Sigma = 342,24$

Reading

Deviazione standard Gruppo B (34 candidati)

Punteggi su base 25	f	fx	x-m	(x-M) ²	f (x-M) ²
20	1	20	6,45	41,60	41,60
19	4	76	5,45	29,70	118,81
18	3	54	4,45	19,80	59,41
16	1	16	2,45	6,00	6,00
15	5	75	1,45	2,10	10,51
14	5	70	0,45	0,20	1,01
13	3	39	-0,55	0,30	0,91
12	3	36	-1,55	2,40	7,21
11	1	11	-2,55	6,50	6,50
10	1	10	-3,55	12,60	12,60
9	3	27	-4,55	20,70	62,11
8	2	16	-5,55	30,80	61,61
6	1	6	-7,55	57,00	57,00
5	1	5	-8,55	73,10	73,10
	$\Sigma = 34$	M = 13,55			$\Sigma = 518,39$

Listening
Punteggi Z e T Gruppo B (34 candidati)

Graduatoria	Candidato	Punteggio su base 20	Punteggi Z	Punteggi T
4	Candidato n. 38	20	1,39	63,93
4	Candidato n. 40	20	1,39	63,93
4	Candidato n. 45	20	1,39	63,93
4	Candidato n. 46	20	1,39	63,93
4	Candidato n. 50	20	1,39	63,93
4	Candidato n. 51	20	1,39	63,93
4	Candidato n. 54	20	1,39	63,93
9	Candidato n. 42	18	0,76	57,63
9	Candidato n. 48	18	0,76	57,63
9	Candidato n. 63	18	0,76	57,63
16	Candidato n. 36	16	0,13	51,32
16	Candidato n. 39	16	0,13	51,32
16	Candidato n. 43	16	0,13	51,32
16	Candidato n. 52	16	0,13	51,32
16	Candidato n. 56	16	0,13	51,32
16	Candidato n. 58	16	0,13	51,32
16	Candidato n. 61	16	0,13	51,32
16	Candidato n. 65	16	0,13	51,32
16	Candidato n. 67	16	0,13	51,32
16	Candidato n. 68	16	0,13	51,32
16	Candidato n. 69	16	0,13	51,32
24,5	Candidato n. 37	14	-0,50	45,02
24,5	Candidato n. 47	14	-0,50	45,02
24,5	Candidato n. 55	14	-0,50	45,02
24,5	Candidato n. 57	14	-0,50	45,02
24,5	Candidato n. 59	14	-0,50	45,02
24,5	Candidato n. 60	14	-0,50	45,02
29,5	Candidato n. 41	12	-1,13	38,72
29,5	Candidato n. 44	12	-1,13	38,72
29,5	Candidato n. 64	12	-1,13	38,72
29,5	Candidato n. 66	12	-1,13	38,72
32,5	Candidato n. 53	10	-1,76	32,41
32,5	Candidato n. 62	10	-1,76	32,41
34	Candidato n. 49	8	-2,39	26,11

Reading
Punteggi Z e T Gruppo B (34 candidati)

Graduatoria	Candidato	Punteggio su base 25	Punteggi Z	Punteggi t
1	Candidato n. 61	20	1,65	66,52
3,5	Candidato n. 36	19	1,40	63,96
3,5	Candidato n. 62	19	1,40	63,96
3,5	Candidato n. 64	19	1,40	63,96
3,5	Candidato n. 69	19	1,40	63,96
7	Candidato n. 39	18	1,14	61,40
7	Candidato n. 59	18	1,14	61,40
7	Candidato n. 60	18	1,14	61,40
9	Candidato n. 52	16	0,63	56,27
12	Candidato n. 43	15	0,37	53,71
12	Candidato n. 46	15	0,37	53,71
12	Candidato n. 50	15	0,37	53,71
12	Candidato n. 58	15	0,37	53,71
12	Candidato n. 66	15	0,37	53,71
17	Candidato n. 38	14	0,12	51,15
17	Candidato n. 40	14	0,12	51,15
17	Candidato n. 54	14	0,12	51,15
17	Candidato n. 67	14	0,12	51,15
17	Candidato n. 68	14	0,12	51,15
21	Candidato n. 37	13	-0,14	48,59
21	Candidato n. 45	13	-0,14	48,59
21	Candidato n. 51	13	-0,14	48,59
24	Candidato n. 41	12	-0,40	46,03
24	Candidato n. 63	12	-0,40	46,03
24	Candidato n. 65	12	-0,40	46,03
26	Candidato n. 42	11	-0,65	43,47
27	Candidato n. 55	10	-0,91	40,91
29	Candidato n. 47	9	-1,17	38,35
29	Candidato n. 53	9	-1,17	38,35
29	Candidato n. 56	9	-1,17	38,35
31,5	Candidato n. 44	8	-1,42	35,79
31,5	Candidato n. 48	8	-1,42	35,79
33	Candidato n. 57	6	-1,93	30,66
34	Candidato n. 49	5	-2,19	28,10

Listening e Reading
Differenze di posizione Gruppo B (34 candidati)

Candidato	Ascolto		Lettura		D	D ²
	Graduatoria	Punteggio T	Graduatoria	Punteggi T		
Candidato n. 36	16	51,32	3,5	63,96	12,5	156,25
Candidato n. 37	24,5	45,02	21	48,59	3,5	12,25
Candidato n. 38	4	63,93	17	51,15	-13	169
Candidato n. 39	16	51,32	7	61,4	9	81
Candidato n. 40	4	63,93	17	51,15	-13	169
Candidato n. 41	29,5	38,72	24	46,03	5,5	30,25
Candidato n. 42	9	57,63	26	43,47	-17	289
Candidato n. 43	16	51,32	12	53,71	4	16
Candidato n. 44	29,5	38,72	31,5	35,79	-2	4
Candidato n. 45	4	63,93	21	48,59	-17	289
Candidato n. 46	4	63,93	12	53,71	-8	64
Candidato n. 47	24,5	45,02	29	38,35	-4,5	20,25
Candidato n. 48	9	57,63	31,5	35,79	-22,5	506,25
Candidato n. 49	34	26,11	34	28,1	0	0
Candidato n. 50	4	63,93	12	53,71	-8	64
Candidato n. 51	4	63,93	21	48,59	-17	289
Candidato n. 52	16	51,32	9	56,27	7	49
Candidato n. 53	32,5	32,41	29	38,35	3,5	12,25
Candidato n. 54	4	63,93	17	51,15	-13	169
Candidato n. 55	24,5	45,02	27	40,91	-2,5	6,25
Candidato n. 56	16	51,32	29	38,35	-13	169
Candidato n. 57	24,5	45,02	33	30,66	-8,5	72,25
Candidato n. 58	16	51,32	12	53,71	4	16
Candidato n. 59	24,5	45,02	7	61,4	17,5	306,25
Candidato n. 60	24,5	45,02	7	61,4	17,5	306,25
Candidato n. 61	16	51,32	1	66,52	15	225
Candidato n. 62	32,5	32,41	3,5	63,96	29	841
Candidato n. 63	9	57,63	24	46,03	-15	225
Candidato n. 64	29,5	38,72	3,5	63,96	26	676
Candidato n. 65	16	51,32	24	46,03	-8	64
Candidato n. 66	29,5	38,72	12	53,71	17,5	306,25
Candidato n. 67	16	51,32	17	51,15	-1	1
Candidato n. 68	16	51,32	17	51,15	-1	1
Candidato n. 69	16	51,32	3,5	63,96	12,5	156,25
						Σ = 5761

Esiti prova di Listening e Reading n. 3

Gruppo C (34 candidati)

Candidato	Prova di ascolto Punteggio su base 20	Prova di lettura Punteggio su base 25
Candidato n. 70	20	15
Candidato n. 71	18	16
Candidato n. 72	20	19
Candidato n. 73	20	16
Candidato n. 74	18	14
Candidato n. 75	20	16
Candidato n. 76	20	13
Candidato n. 77	20	8
Candidato n. 78	18	19
Candidato n. 79	12	19
Candidato n. 80	16	17
Candidato n. 81	16	17
Candidato n. 82	16	17
Candidato n. 83	18	14
Candidato n. 84	18	14
Candidato n. 85	20	22
Candidato n. 86	20	20
Candidato n. 87	20	20
Candidato n. 88	18	13
Candidato n. 89	20	14
Candidato n. 90	20	20
Candidato n. 91	20	20
Candidato n. 92	20	20
Candidato n. 93	18	20
Candidato n. 94	18	19
Candidato n. 95	18	20
Candidato n. 96	18	20
Candidato n. 97	20	19
Candidato n. 98	20	22
Candidato n. 99	20	19
Candidato n. 100	20	20
Candidato n. 101	20	19
Candidato n. 102	20	17
Candidato n. 103	16	20

Listening

Gruppo C *higher* (9 candidati):

Candidato	Prova di ascolto (punteggio su base 20)
Candidato n. 70	20
Candidato n. 72	20
Candidato n. 73	20
Candidato n. 75	20
Candidato n. 76	20
Candidato n. 77	20
Candidato n. 85	20
Candidato n. 86	20
Candidato n. 87	20

Gruppo C *lower* (9 candidati):

Candidato	Prova di ascolto (punteggio su base 20)
Candidato n. 93	18
Candidato n. 94	18
Candidato n. 95	18
Candidato n. 96	18
Candidato n. 80	16
Candidato n. 81	16
Candidato n. 82	16
Candidato n. 94	16
Candidato n. 79	12

Reading

Gruppo C *higher* (9 candidati):

Candidato	Prova di lettura (punteggio su base 25)
Candidato n. 85	22
Candidato n. 98	22
Candidato n. 86	20
Candidato n. 87	20
Candidato n. 90	20
Candidato n. 91	20
Candidato n. 92	20
Candidato n. 93	20
Candidato n. 95	20

Gruppo C *lower* (9 candidati):

Candidato	Prova di lettura (punteggio su base 25)
Candidato n. 75	16
Candidato n. 70	15
Candidato n. 74	14
Candidato n. 83	14
Candidato n. 84	14
Candidato n. 89	14
Candidato n. 76	13
Candidato n. 88	13
Candidato n. 77	8

Listening

Deviazione standard Gruppo C (34 candidati)

Punteggi su base 20	f	fx	x-m	(x-M) ²	f (x-M) ²
20	19	380	1,30	1,69	32,11
18	10	180	-0,70	0,49	4,90
16	4	64	-2,70	7,29	29,16
12	1	12	-6,70	44,89	44,89
	$\Sigma = 34$	M = 18,70			$\Sigma = 111,06$

Reading

Deviazione standard Gruppo C (34 candidati)

Punteggi su base 25	f	fx	x-m	(x-M) ²	f (x-M) ²
22	2	44	4,42	19,54	39,07
20	10	200	2,42	5,86	58,56
19	7	133	1,42	2,02	14,11
17	4	68	-0,58	0,34	1,35
16	3	48	-1,58	2,50	7,49
15	1	15	-2,58	6,66	6,66
14	4	56	-3,58	12,82	51,27
13	2	26	-4,58	20,98	41,95
8	1	8	-9,58	91,78	91,78
	$\Sigma = 34$	M = 17,58			$\Sigma = 312,24$

Listening
Punteggi Z e T Gruppo C (34 candidati)

Graduatoria	Candidato	Punteggio su base 20	Punteggi Z	Punteggi T
10	Candidato n. 70	20	0,72	57,19
10	Candidato n. 72	20	0,72	57,19
10	Candidato n. 73	20	0,72	57,19
10	Candidato n. 75	20	0,72	57,19
10	Candidato n. 76	20	0,72	57,19
10	Candidato n. 77	20	0,72	57,19
10	Candidato n. 85	20	0,72	57,19
10	Candidato n. 86	20	0,72	57,19
10	Candidato n. 87	20	0,72	57,19
10	Candidato n. 89	20	0,72	57,19
10	Candidato n. 90	20	0,72	57,19
10	Candidato n. 91	20	0,72	57,19
10	Candidato n. 92	20	0,72	57,19
10	Candidato n. 97	20	0,72	57,19
10	Candidato n. 98	20	0,72	57,19
10	Candidato n. 99	20	0,72	57,19
10	Candidato n. 100	20	0,72	57,19
10	Candidato n. 101	20	0,72	57,19
10	Candidato n. 102	20	0,72	57,19
24,5	Candidato n. 71	18	-0,39	46,13
24,5	Candidato n. 74	18	-0,39	46,13
24,5	Candidato n. 78	18	-0,39	46,13
24,5	Candidato n. 83	18	-0,39	46,13
24,5	Candidato n. 84	18	-0,39	46,13
24,5	Candidato n. 88	18	-0,39	46,13
24,5	Candidato n. 93	18	-0,39	46,13
24,5	Candidato n. 94	18	-0,39	46,13
24,5	Candidato n. 95	18	-0,39	46,13
24,5	Candidato n. 96	18	-0,39	46,13
31,5	Candidato n. 80	16	-1,49	35,06
31,5	Candidato n. 81	16	-1,49	35,06
31,5	Candidato n. 82	16	-1,49	35,06
31,5	Candidato n. 103	16	-1,49	35,06
34	Candidato n. 79	12	-3,71	12,93

Reading
Punteggi Z e T Gruppo C (34 candidati)

Graduatoria	Candidato	Punteggio su base 25	Punteggi Z	Punteggi T
1,5	Candidato n. 85	22	1,46	64,59
1,5	Candidato n. 98	22	1,46	64,59
7,5	Candidato n. 86	20	0,80	57,99
7,5	Candidato n. 87	20	0,80	57,99
7,5	Candidato n. 90	20	0,80	57,99
7,5	Candidato n. 91	20	0,80	57,99
7,5	Candidato n. 92	20	0,80	57,99
7,5	Candidato n. 93	20	0,80	57,99
7,5	Candidato n. 95	20	0,80	57,99
7,5	Candidato n. 96	20	0,80	57,99
7,5	Candidato n. 100	20	0,80	57,99
7,5	Candidato n. 103	20	0,80	57,99
16	Candidato n. 72	19	0,47	54,69
16	Candidato n. 78	19	0,47	54,69
16	Candidato n. 79	19	0,47	54,69
16	Candidato n. 94	19	0,47	54,69
16	Candidato n. 97	19	0,47	54,69
16	Candidato n. 99	19	0,47	54,69
16	Candidato n. 101	19	0,47	54,69
21,5	Candidato n. 80	17	-0,19	48,09
21,5	Candidato n. 81	17	-0,19	48,09
21,5	Candidato n. 82	17	-0,19	48,09
21,5	Candidato n. 102	17	-0,19	48,09
25	Candidato n. 71	16	-0,52	44,79
25	Candidato n. 73	16	-0,52	44,79
25	Candidato n. 75	16	-0,52	44,79
27	Candidato n. 70	15	-0,85	41,49
29,5	Candidato n. 74	14	-1,18	38,19
29,5	Candidato n. 83	14	-1,18	38,19
29,5	Candidato n. 84	14	-1,18	38,19
29,5	Candidato n. 89	14	-1,18	38,19
32,5	Candidato n. 76	13	-1,51	34,89
32,5	Candidato n. 88	13	-1,51	34,89
34	Candidato n. 77	8	-3,16	18,39

Listening e Reading
Differenze di posizione Gruppo C (34 candidati)

Candidato	Ascolto		Lettura		D	D ²
	Graduatoria	Punteggio T	Graduatoria	Punteggi T		
Candidato n. 70	10	57,19	27	41,49	-17	289
Candidato n. 71	24,5	46,13	25	44,79	-0,5	0,25
Candidato n. 72	10	57,19	16	54,69	-6	36
Candidato n. 73	10	57,19	25	44,79	-15	225
Candidato n. 74	24,5	46,13	29,5	38,19	-5	25
Candidato n. 75	10	57,19	25	44,79	-15	225
Candidato n. 76	10	57,19	32,5	34,89	-22,5	506,25
Candidato n. 77	10	57,19	34	18,39	-24	576
Candidato n. 78	24,5	46,13	16	54,69	8,5	72,25
Candidato n. 79	34	12,93	16	54,69	18	324
Candidato n. 80	31,5	35,06	21,5	48,09	10	100
Candidato n. 81	31,5	35,06	21,5	48,09	10	100
Candidato n. 82	31,5	35,06	21,5	48,09	10	100
Candidato n. 83	24,5	46,13	29,5	38,19	-5	25
Candidato n. 84	24,5	46,13	29,5	38,19	-5	25
Candidato n. 85	10	57,19	1,5	64,59	8,5	72,25
Candidato n. 86	10	57,19	7,5	57,99	2,5	6,25
Candidato n. 87	10	57,19	7,5	57,99	2,5	6,25
Candidato n. 88	24,5	46,13	32,5	34,89	-8	64
Candidato n. 89	10	57,19	29,5	38,19	-19,5	380,25
Candidato n. 90	10	57,19	7,5	57,99	2,5	6,25
Candidato n. 91	10	57,19	7,5	57,99	2,5	6,25
Candidato n. 92	10	57,19	7,5	57,99	2,5	6,25
Candidato n. 93	24,5	46,13	7,5	57,99	17	289
Candidato n. 94	24,5	46,13	16	54,69	8,5	72,25
Candidato n. 95	24,5	46,13	7,5	57,99	17	289
Candidato n. 96	24,5	46,13	7,5	57,99	17	289
Candidato n. 97	10	57,19	16	54,69	-6	36
Candidato n. 98	10	57,19	1,5	64,59	8,5	72,25
Candidato n. 99	10	57,19	16	54,69	-6	36
Candidato n. 100	10	57,19	7,5	57,99	2,5	6,25
Candidato n. 101	10	57,19	16	54,69	-6	36
Candidato n. 102	10	57,19	21,5	48,09	-11,5	132,25
Candidato n. 103	31,5	35,06	7,5	57,99	24	576
						Σ = 5010,5

Esiti prova di Listening e Reading n. 4

Gruppo D (34 candidati)

Candidato	Prova di ascolto Punteggio su base 20	Prova di lettura Punteggio su base 25
Candidato n. 104	18	23
Candidato n. 105	20	23
Candidato n. 106	14	17
Candidato n. 107	12	18
Candidato n. 108	16	22
Candidato n. 109	20	22
Candidato n. 110	18	20
Candidato n. 111	18	14
Candidato n. 112	18	22
Candidato n. 113	12	15
Candidato n. 114	18	15
Candidato n. 115	20	24
Candidato n. 116	20	21
Candidato n. 117	20	22
Candidato n. 118	16	14
Candidato n. 119	20	23
Candidato n. 120	18	19
Candidato n. 121	20	23
Candidato n. 122	18	21
Candidato n. 123	20	13
Candidato n. 124	18	21
Candidato n. 125	16	19
Candidato n. 126	16	13
Candidato n. 127	16	16
Candidato n. 128	14	17
Candidato n. 129	18	24
Candidato n. 130	16	13
Candidato n. 131	16	17
Candidato n. 132	14	19
Candidato n. 133	20	15
Candidato n. 134	18	19
Candidato n. 135	18	19
Candidato n. 136	18	15
Candidato n. 137	18	20

Listening

Gruppo D *higher* (9 candidati):

Candidato	Prova di ascolto (punteggio su base 20)
Candidato n. 105	20
Candidato n. 109	20
Candidato n. 115	20
Candidato n. 116	20
Candidato n. 117	20
Candidato n. 119	20
Candidato n. 121	20
Candidato n. 123	20
Candidato n. 133	20

Gruppo D *lower* (9 candidati):

Candidato	Prova di ascolto (punteggio su base 20)
Candidato n. 126	16
Candidato n. 127	16
Candidato n. 130	16
Candidato n. 131	16
Candidato n. 106	14
Candidato n. 128	14
Candidato n. 132	14
Candidato n. 107	12
Candidato n. 113	12

Reading

Gruppo D *higher* (9 candidati):

Candidato	Prova di lettura (punteggio su base 25)
Candidato n. 115	24
Candidato n. 129	24
Candidato n. 104	23
Candidato n. 105	23
Candidato n. 119	23
Candidato n. 121	23
Candidato n. 108	22
Candidato n. 109	22
Candidato n. 112	22

Gruppo D *lower* (9 candidati):

Candidato	Prova di lettura (punteggio su base 25)
Candidato n. 113	15
Candidato n. 114	15
Candidato n. 133	15
Candidato n. 136	15
Candidato n. 111	14
Candidato n. 118	14
Candidato n. 123	13
Candidato n. 126	13
Candidato n. 130	13

Listening

Deviazione standard Gruppo D (34 candidati)

Punteggi su base 20	f	fx	x-m	(x-M) ²	f (x-M) ²
20	9	180	2,59	6,71	60,37
18	13	234	0,59	0,35	4,53
16	7	112	-1,41	1,99	13,92
14	3	42	-3,41	11,63	34,88
12	2	24	-5,41	29,27	58,54
	$\Sigma = 34$	M = 17,41			$\Sigma = 172,24$

Reading

Deviazione standard Gruppo D (34 candidati)

Punteggi su base 25	f	fx	x-m	(x-M) ²	f (x-M) ²
24	2	48	5,24	27,46	54,92
23	4	92	4,24	17,98	71,91
22	4	88	3,24	10,50	41,99
21	3	63	2,24	5,02	15,05
20	2	40	1,24	1,54	3,08
19	5	95	0,24	0,06	0,29
18	1	18	-0,76	0,58	0,58
17	3	51	-1,76	3,10	9,29
16	1	16	-2,76	7,62	7,62
15	4	60	-3,76	14,14	56,55
14	2	28	-4,76	22,66	45,32
13	3	39	-5,76	33,18	99,53
	$\Sigma = 34$	M = 18,76			$\Sigma = 406,12$

Listening
Punteggi Z e T Gruppo D (34 candidati)

Graduatoria	Candidato	Punteggio su base 20	Punteggi Z	Punteggi T
5	Candidato n. 105	20	1,15	61,51
5	Candidato n. 109	20	1,15	61,51
5	Candidato n. 115	20	1,15	61,51
5	Candidato n. 116	20	1,15	61,51
5	Candidato n. 117	20	1,15	61,51
5	Candidato n. 119	20	1,15	61,51
5	Candidato n. 121	20	1,15	61,51
5	Candidato n. 123	20	1,15	61,51
5	Candidato n. 133	20	1,15	61,51
16	Candidato n. 104	18	0,26	52,62
16	Candidato n. 110	18	0,26	52,62
16	Candidato n. 111	18	0,26	52,62
16	Candidato n. 112	18	0,26	52,62
16	Candidato n. 114	18	0,26	52,62
16	Candidato n. 120	18	0,26	52,62
16	Candidato n. 122	18	0,26	52,62
16	Candidato n. 124	18	0,26	52,62
16	Candidato n. 129	18	0,26	52,62
16	Candidato n. 134	18	0,26	52,62
16	Candidato n. 135	18	0,26	52,62
16	Candidato n. 136	18	0,26	52,62
16	Candidato n. 137	18	0,26	52,62
26	Candidato n. 108	16	-0,63	43,74
26	Candidato n. 118	16	-0,63	43,74
26	Candidato n. 125	16	-0,63	43,74
26	Candidato n. 126	16	-0,63	43,74
26	Candidato n. 127	16	-0,63	43,74
26	Candidato n. 130	16	-0,63	43,74
26	Candidato n. 131	16	-0,63	43,74
31	Candidato n. 106	14	-1,52	34,85
31	Candidato n. 128	14	-1,52	34,85
31	Candidato n. 132	14	-1,52	34,85
33,5	Candidato n. 107	12	-2,4	25,96
33,5	Candidato n. 113	12	-2,4	25,96

Reading
Punteggi Z e T Gruppo D (34 candidati)

Graduatoria	Candidato	Punteggio su base 25	Punteggi Z	Punteggi T
1,5	Candidato n. 115	24	1,52	65,16
1,5	Candidato n. 129	24	1,52	65,16
4,5	Candidato n. 104	23	1,23	62,27
4,5	Candidato n. 105	23	1,23	62,27
4,5	Candidato n. 119	23	1,23	62,27
4,5	Candidato n. 121	23	1,23	62,27
8,5	Candidato n. 108	22	0,94	59,37
8,5	Candidato n. 109	22	0,94	59,37
8,5	Candidato n. 112	22	0,94	59,37
8,5	Candidato n. 117	22	0,94	59,37
12	Candidato n. 116	21	0,65	56,48
12	Candidato n. 122	21	0,65	56,48
12	Candidato n. 124	21	0,65	56,48
14,5	Candidato n. 110	20	0,36	53,59
14,5	Candidato n. 137	20	0,36	53,59
18	Candidato n. 120	19	0,07	50,69
18	Candidato n. 125	19	0,07	50,69
18	Candidato n. 132	19	0,07	50,69
18	Candidato n. 134	19	0,07	50,69
18	Candidato n. 135	19	0,07	50,69
21	Candidato n. 107	18	-0,22	47,80
23	Candidato n. 106	17	-0,51	44,91
23	Candidato n. 128	17	-0,51	44,91
23	Candidato n. 131	17	-0,51	44,91
25	Candidato n. 127	16	-0,80	42,01
27,5	Candidato n. 113	15	-1,09	39,12
27,5	Candidato n. 114	15	-1,09	39,12
27,5	Candidato n. 133	15	-1,09	39,12
27,5	Candidato n. 136	15	-1,09	39,12
30,5	Candidato n. 111	14	-1,38	36,23
30,5	Candidato n. 118	14	-1,38	36,23
33	Candidato n. 123	13	-1,67	33,33
33	Candidato n. 126	13	-1,67	33,33
33	Candidato n. 130	13	-1,67	33,33

Listening e Reading
Differenze di posizione Gruppo D (34 candidati)

Candidato	Ascolto		Lettura		D	D ²
	Graduatoria	Punteggio T	Graduatoria	Punteggi T		
Candidato n. 104	16	52,62	4,5	62,27	11,5	132,25
Candidato n. 105	5	61,51	4,5	62,27	0,5	0,25
Candidato n. 106	31	34,85	23	44,91	8	64
Candidato n. 107	33,5	25,96	21	47,8	12,5	156,25
Candidato n. 108	26	43,74	8,5	59,37	17,5	306,25
Candidato n. 109	5	61,51	8,5	59,37	-3,5	12,25
Candidato n. 110	16	52,62	14,5	53,59	1,5	2,25
Candidato n. 111	16	52,62	30,5	36,23	-14,5	210,25
Candidato n. 112	16	52,62	8,5	59,37	7,5	56,25
Candidato n. 113	33,5	25,96	27,5	39,12	6	36
Candidato n. 114	16	52,62	27,5	39,12	-11,5	132,25
Candidato n. 115	5	61,51	1,5	65,16	3,5	12,25
Candidato n. 116	5	61,51	12	56,48	-7	49
Candidato n. 117	5	61,51	8,5	59,37	-3,5	12,25
Candidato n. 118	26	43,74	30,5	36,23	-4,5	20,25
Candidato n. 119	5	61,51	4,5	62,27	0,5	0,25
Candidato n. 120	16	52,62	18	50,69	-2	4
Candidato n. 121	5	61,51	4,5	62,27	0,5	0,25
Candidato n. 122	16	52,62	12	56,48	4	16
Candidato n. 123	5	61,51	33	33,33	-28	784
Candidato n. 124	16	52,62	12	56,48	4	16
Candidato n. 125	26	43,74	18	50,69	8	64
Candidato n. 126	26	43,74	33	33,33	-7	49
Candidato n. 127	26	43,74	25	42,01	1	1
Candidato n. 128	31	34,85	23	44,91	8	64
Candidato n. 129	16	52,62	1,5	65,16	14,5	210,25
Candidato n. 130	26	43,74	33	33,33	-7	49
Candidato n. 131	26	43,74	23	44,91	3	9
Candidato n. 132	31	34,85	18	50,69	13	169
Candidato n. 133	5	61,51	27,5	39,12	-22,5	506,25
Candidato n. 134	16	52,62	18	50,69	-2	4
Candidato n. 135	16	52,62	18	50,69	-2	4
Candidato n. 136	16	52,62	27,5	39,12	-11,5	132,25
Candidato n. 137	16	52,62	14,5	53,59	1,5	2,25
						Σ = 3286,5

Esiti prova di Listening e Reading n. 5

Gruppo E (17 candidati)

Candidato	Prova di ascolto Punteggio su base 20	Prova di lettura Punteggio su base 25
Candidato n. 138	12	16
Candidato n. 139	6	8
Candidato n. 140	6	7
Candidato n. 141	6	10
Candidato n. 142	6	13
Candidato n. 143	6	12
Candidato n. 144	6	11
Candidato n. 145	6	10
Candidato n. 146	6	10
Candidato n. 147	6	12
Candidato n. 148	8	15
Candidato n. 149	12	6
Candidato n. 150	10	6
Candidato n. 151	10	5
Candidato n. 152	8	19
Candidato n. 153	8	19
Candidato n. 154	12	24

Listening

Gruppo E *higher* (5 candidati):

Candidato	Prova di ascolto (punteggio su base 20)
Candidato n. 138	12
Candidato n. 149	12
Candidato n. 154	12
Candidato n. 150	10
Candidato n. 151	10

Gruppo E *lower* (5 candidati):

Candidato	Prova di ascolto (punteggio su base 20)
Candidato n. 143	6
Candidato n. 144	6
Candidato n. 145	6
Candidato n. 146	6
Candidato n. 147	6

Reading

Gruppo E *higher* (5 candidati):

Candidato	Prova di lettura (punteggio su base 25)
Candidato n. 154	24
Candidato n. 152	19
Candidato n. 153	19
Candidato n. 138	16
Candidato n. 148	15

Gruppo E *lower* (5 candidati):

Candidato	Prova di lettura (punteggio su base 25)
Candidato n. 139	8
Candidato n. 140	7
Candidato n. 149	6
Candidato n. 150	6
Candidato n. 151	5

Listening

Deviazione standard Gruppo E (17 candidati)

Punteggi su base 20	f	fx	x-m	(x-M) ²	f (x-M) ²
12	3	36	4,12	16,97	50,92
10	2	20	2,12	4,49	8,99
8	3	24	0,12	0,01	0,04
6	9	54	-1,88	3,53	31,81
	$\Sigma = 17$	M = 7,88			$\Sigma = 91,76$

Reading

Deviazione standard Gruppo E (17 candidati)

Punteggi su base 25	f	fx	x-m	(x-M) ²	f (x-M) ²
24	1	24	12,06	145,44	145,44
19	2	38	7,06	49,84	99,69
16	1	16	4,06	16,48	16,48
15	1	15	3,06	9,36	9,36
13	1	13	1,06	1,12	1,12
12	2	24	0,06	0,00	0,01
11	1	11	-0,94	0,88	0,88
10	3	30	-1,94	3,76	11,29
8	1	8	-3,94	15,52	15,52
7	1	7	-4,94	24,40	24,40
6	2	12	-5,94	35,28	70,57
5	1	5	-6,94	48,16	48,16
	$\Sigma = 17$	M = 11,94			$\Sigma = 442,94$

Listening
Punteggi Z e T Gruppo E (17 candidati)

Graduatoria	Candidato	Punteggio su base 20	Punteggi Z	Punteggi T
2	Candidato n. 138	12	1,77	67,73
2	Candidato n. 149	12	1,77	67,73
2	Candidato n. 154	12	1,77	67,73
4,5	Candidato n. 150	10	0,91	59,12
4,5	Candidato n. 151	10	0,91	59,12
7	Candidato n. 148	8	0,05	50,52
7	Candidato n. 152	8	0,05	50,52
7	Candidato n. 153	8	0,05	50,52
13	Candidato n. 139	6	-0,81	41,91
13	Candidato n. 140	6	-0,81	41,91
13	Candidato n. 141	6	-0,81	41,91
13	Candidato n. 142	6	-0,81	41,91
13	Candidato n. 143	6	-0,81	41,91
13	Candidato n. 144	6	-0,81	41,91
13	Candidato n. 145	6	-0,81	41,91
13	Candidato n. 146	6	-0,81	41,91
13	Candidato n. 147	6	-0,81	41,91

Reading
Punteggi Z e T Gruppo E (17 candidati)

Graduatoria	Candidato	Punteggio su base 25	Punteggi Z	Punteggi T
1	Candidato n. 154	24	2,36	73,63
2,5	Candidato n. 152	19	1,38	63,83
2,5	Candidato n. 153	19	1,38	63,83
4	Candidato n. 138	16	0,80	57,95
5	Candidato n. 148	15	0,60	55,99
6	Candidato n. 142	13	0,21	52,08
7,5	Candidato n. 143	12	0,01	50,12
7,5	Candidato n. 147	12	0,01	50,12
9	Candidato n. 144	11	-0,18	48,16
11	Candidato n. 141	10	-0,38	46,20
11	Candidato n. 145	10	-0,38	46,20
11	Candidato n. 146	10	-0,38	46,20
13	Candidato n. 139	8	-0,77	42,28
14	Candidato n. 140	7	-0,97	40,32
15,5	Candidato n. 149	6	-1,16	38,36
15,5	Candidato n. 150	6	-1,16	38,36
17	Candidato n. 151	5	-1,36	36,40

Listening e Reading
Differenze di posizione Gruppo E (17 candidati)

Candidato	Ascolto		Lettura		D	D ²
	Graduatoria	Punteggio T	Graduatoria	Punteggi T		
Candidato n. 138	2	67,73	4	57,95	-2	4
Candidato n. 139	13	41,91	13	42,28	0	0
Candidato n. 140	13	41,91	14	40,32	-1	1
Candidato n. 141	13	41,91	11	46,2	2	4
Candidato n. 142	13	41,91	6	52,08	7	49
Candidato n. 143	13	41,91	7,5	50,12	5,5	30,25
Candidato n. 144	13	41,91	9	48,16	4	16
Candidato n. 145	13	41,91	11	46,2	2	4
Candidato n. 146	13	41,91	11	46,2	2	4
Candidato n. 147	13	41,91	7,5	50,12	5,5	30,25
Candidato n. 148	7	50,52	5	55,99	2	4
Candidato n. 149	2	67,73	15,5	38,36	-13,5	182,25
Candidato n. 150	4,5	59,12	15,5	38,36	-11	121
Candidato n. 151	4,5	59,12	17	36,4	-12,5	156,25
Candidato n. 152	7	50,52	2,5	63,83	4,5	20,25
Candidato n. 153	7	50,52	2,5	63,83	4,5	20,25
Candidato n. 154	2	67,73	1	73,63	1	1
						Σ = 647,5

Item Tally Sheet – Prova di Listening n. 1

Tipo di prova: ascolto		Gruppo: A	Totale candidati: 20		Data: 31 marzo 2009		
N. item	Risp	Candidati gruppo A _n : 10	p	Candidati gruppo A _i : 10	p	IF	ID
Parte prima							
1	a	-	0	-	0	0	0
	b	10	1	9	0,9	0,95	0,1
	c	-	0	1	0,1	0,05	-0,1
	O	-	0	-	0	0	0
2	a	10	1	10	1	1	0
	b	-	0	-	0	0	0
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
3	a	-	0	-	0	0	0
	b	-	0	-	0	0	0
	c	10	1	10	1	1	0
	O	-	0	-	0	0	0
4	a	-	0	4	0,4	0,2	-0,4
	b	-	0	2	0,2	0,1	-0,2
	c	10	1	4	0,4	0,7	0,6
	O	-	0	-	0	0	0
5	a	-	0	-	0	0	0
	b	10	1	10	1	1	0
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
Parte seconda							
1	a	-	0	-	0	0	0
	b	10	1	10	1	1	0
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
2	a	-	0	6	0,6	0,3	-0,6
	b	-	0	2	0,2	0,1	-0,2
	c	10	1	2	0,2	0,6	0,8
	O	-	0	-	0	0	0
3	a	-	0	-	0	0	0
	b	10	1	10	1	1	0
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
4	a	7	0,7	10	1	0,85	-0,3
	b	-	0	-	0	0	0
	c	3	0,3	0	0	0,15	0,3
	O	-	0	-	0	0	0
5	a	-	0	-	0	0	0
	b	3	0,3	6	0,6	0,45	-0,3
	c	7	0,7	4	0,4	0,55	0,3
	O	-	0	-	0	0	0

Item Tally Sheet – Prova di Listening n. 2

Tipo di prova: ascolto		Gruppo: B		Totale candidati: 18		Data: 31 marzo 2009	
N. item	Risp	Candidati gruppo B _n : 9	p	Candidati gruppo B _i : 9	p	IF	ID
Parte prima							
1	a	9	1	6	0,67	0,84	0,34
	b	-	0	-	0	0	0
	c	-	0	3	0,34	0,17	-0,34
	O	-	0	-	0	0	0
2	a	-	0	-	0	0	0
	b	9	1	6	0,67	0,84	0,34
	c	-	0	3	0,34	0,17	-0,34
	O	-	0	-	0	0	0
3	a	-	0	-	0	0	0
	b	9	1	9	1	1	0
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
4	a	-	0	3	0,34	0,17	-0,34
	b	-	0	-	0	0	0
	c	9	1	6	0,67	0,84	0,34
	O	-	0	-	0	0	0
5	a	-	0	-	0	0	0
	b	-	0	4	0,45	0,23	-0,45
	c	9	1	5	0,56	0,78	0,45
	O	-	0	-	0	0	0
Parte seconda							
1	a	9	1	8	0,89	0,95	0,12
	b	-	0	-	0	0	0
	c	-	0	1	0,12	0,06	-0,12
	O	-	0	-	0	0	0
2	a	-	0	-	0	0	0
	b	1	0,12	1	0,12	0,12	0
	c	8	0,89	8	0,89	0,89	0
	O	-	0	-	0	0	0
3	a	-	0	3	0,34	0,17	-0,34
	b	-	0	-	0	0	0
	c	9	1	6	0,67	0,84	0,34
	O	-	0	-	0	0	0
4	a	9	1	3	0,34	0,67	0,67
	b	-	0	3	0,34	0,17	-0,34
	c	-	0	3	0,34	0,17	-0,34
	O	-	0	-	0	0	0
5	a	-	0	3	0,34	0,17	-0,34
	b	8	0,89	-	0	0,45	0,89
	c	1	0,12	6	0,67	0,39	-0,56
	O	-	0	-	0	0	0

Item Tally Sheet – Prova di Listening n. 3

Tipo di prova: ascolto		Gruppo: C		Totale candidati: 18		Data: 31 marzo 2009	
N. item	Risp	Candidati gruppo C _n : 9	p	Candidati gruppo C _i : 9	p	IF	ID
Parte prima							
1	a	-	0	-	0	0	0
	b	9	1	7	0,78	0,89	0,23
	c	-	0	2	0,23	0,12	-0,23
	O	-	0	-	0	0	0
2	a	-	0	-	0	0	0
	b	-	0	-	0	0	0
	c	9	1	9	1	1	0
	O	-	0	-	0	0	0
3	a	-	0	3	0,34	0,17	-0,34
	b	9	1	3	0,34	0,67	0,67
	c	-	0	3	0,34	0,17	-0,34
	O	-	0	-	0	0	0
4	a	9	1	6	0,67	0,84	0,34
	b	-	0	3	0,34	0,17	-0,34
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
5	a	-	0	-	0	0	0
	b	9	1	9	1	1	0
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
Parte seconda							
1	a	-	0	1	0,12	0,06	-0,12
	b	9	1	7	0,78	0,89	0,23
	c	-	0	1	0,12	0,06	-0,12
	O	-	0	-	0	0	0
2	a	-	0	-	0	0	0
	b	-	0	-	0	0	0
	c	9	1	9	1	1	0
	O	-	0	-	0	0	0
3	a	-	0	1	0,12	0,06	-0,12
	b	9	1	8	0,89	0,95	0,12
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
4	a	9	1	7	0,78	0,89	0,23
	b	-	0	-	0	0	0
	c	-	0	2	0,23	0,12	-0,23
	O	-	0	-	0	0	0
5	a	-	0	-	0	0	0
	b	-	0	-	0	0	0
	c	9	1	9	1	1	0
	O	-	0	-	0	0	0

Item Tally Sheet – Prova di Listening n. 4

Tipo di prova: ascolto		Gruppo: D		Totale candidati: 18		Data: 31 marzo 2009	
N. item	Risp	Candidati gruppo D _n : 9	p	Candidati gruppo D _i : 9	p	IF	ID
Parte prima							
1	a	0	0	1	0,12	0,06	-0,12
	b	9	1	7	0,78	0,89	0,23
	c	0	0	1	0,12	0,06	-0,12
	O	-	0	-	0	0	0
2	a	0	0	2	0,23	0,12	-0,23
	b	9	1	7	0,78	0,89	0,23
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
3	a	-	0	-	0	0	0
	b	-	0	-	0	0	0
	c	9	1	9	1	1	0
	O	-	0	-	0	0	0
4	a	-	0	-	0	0	0
	b	9	1	9	1	1	0
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
5	a	-	0	-	0	0	0
	b	-	0	-	0	0	0
	c	9	1	9	1	1	0
	O	-	0	-	0	0	0
Parte seconda							
1	a	-	0	2	0,23	0,12	-0,23
	b	-	0	-	0	0	0
	c	9	1	7	0,78	0,89	0,23
	O	-	0	-	0	0	0
2	a	9	1	6	0,67	0,84	0,34
	b	-	0	3	0,34	0,17	-0,34
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
3	a	-	0	-	0	0	0
	b	-	0	-	0	0	0
	c	9	1	9	1	1	0
	O	-	0	-	0	0	0
4	a	-	0	1	0,12	0,06	-0,12
	b	9	1	8	0,89	0,95	0,12
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
5	a	9	1	2	0,23	0,62	0,78
	b	-	0	1	0,12	0,06	-0,12
	c	-	0	6	0,67	0,34	-0,67
	O	-	0	-	0	0	0

Item Tally Sheet – Prova di Listening n. 5

Tipo di prova: ascolto		Gruppo: E		Totale candidati: 10		Data: 29 giugno 2009	
N. item	Risp	Candidati gruppo E _n : 5	p	Candidati gruppo E _i : 5	p	IF	ID
Parte prima							
1	a	2	0,4	-	0	0,2	0,4
	b	1	0,2	1	0,2	0,2	0
	c	2	0,4	4	0,8	0,6	-0,4
	O	-	0	-	0	0	0
2	a	-	0	1	0,2	0,1	-0,2
	b	1	0,2	3	0,6	0,4	-0,4
	c	4	0,8	1	0,2	0,5	0,6
	O	-	0	-	0	0	0
3	a	-	0	-	0	0	0
	b	-	0	-	0	0	0
	c	5	1	5	1	1	0
	O	-	0	-	0	0	0
4	a	-	0	-	0	0	0
	b	3	0,6	5	1	0,8	-0,4
	c	2	0,4	-	0	0,2	0,4
	O	-	0	-	0	0	0
5	a	5	1	-	0	0,5	1
	b	-	0	5	1	0,5	-1
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
Parte seconda							
1	a	2	0,4	2	0,4	0,4	0
	b	1	0,2	-	0	0,1	0,2
	c	2	0,4	3	0,6	0,5	-0,2
	O	-	0	-	0	0	0
2	a	4	0,8	3	0,6	0,7	0,2
	b	-	0	2	0,4	0,2	-0,4
	c	1	0,2	-	0	0,1	0,2
	O	-	0	-	0	0	0
3	a	1	0,2	3	0,6	0,4	-0,4
	b	3	0,6	2	0,4	0,5	0,2
	c	1	0,2	-	0	0,1	0,2
	O	-	0	-	0	0	0
4	a	-	0	-	0	0	0
	b	3	0,6	1	0,2	0,4	0,4
	c	2	0,4	4	0,8	0,6	-0,4
	O	-	0	-	0	0	0
5	a	-	0	2	0,4	0,2	-0,4
	b	4	0,8	-	0	0,4	0,8
	c	1	0,2	3	0,6	0,4	-0,4
	O	-	0	-	0	0	0

Item Tally Sheet - Prova di Reading n. 1

Tipo di prova: lettura		Gruppo: A		Totale candidati: 20		Data: 31 marzo 2009	
N. item	Risp	Candidati gruppo A _H :10	p	Candidati gruppo A _I :10	p	IF	ID
Parte prima							
1	<u>a</u>	10	1	10	1	1	0
	b	-	0	-	0	0	0
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
2	a	-	0	-	0	0	0
	<u>b</u>	10	1	10	1	1	0
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
3	a	-	0	2	0,2	0,1	-0,2
	b	-	0	-	0	0	0
	<u>c</u>	10	1	8	0,8	0,9	0,2
	O	-	0	-	0	0	0
4	a	-	0	4	0,4	0,2	-0,4
	<u>b</u>	10	1	6	0,6	0,8	0,4
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
5	<u>a</u>	10	1	9	0,9	0,95	0,1
	b	-	0	-	0	0	0
	c	-	0	1	0,1	0,05	-0,1
	O	-	0	-	0	0	0
Parte seconda							
1	a	-	0	-	0	0	0
	b	-	0	2	0,2	0,1	-0,2
	<u>c</u>	9	0,9	8	0,8	0,85	0,1
	O	1	0,1	-	0	0,05	0,1
2	a	-	0	1	0,1	0,05	-0,1
	b	-	0	-	0	0	0
	<u>c</u>	10	1	9	0,9	0,95	0,1
	O	-	0	-	0	0	0
3	a	-	0	4	0,4	0,2	-0,4
	<u>b</u>	10	1	6	0,6	0,8	0,4
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
4	<u>a</u>	9	0,9	-	0	0,45	0,9
	b	1	0,1	7	0,7	0,4	-0,6
	c	-	0	-	0	0	0
	O	-	0	3	0,3	0,15	-0,3
5	a	-	0	-	0	0	0
	<u>b</u>	10	1	4	0,4	0,7	0,6
	c	-	0	4	0,4	0,2	-0,4
	O	-	0	2	0,2	0,1	-0,2
Parte terza							
1	<u>vero</u>	10	1	8	0,8	0,9	0,2
	falso	-	0	1	0,1	0,05	-0,1
	O	-	0	1	0,1	0,05	-0,1
2	<u>vero</u>	4	0,4	3	0,3	0,35	0,1
	<u>falso</u>	6	0,6	6	0,6	0,6	0

	O	-	0	1	0,1	0,05	-0,1
3	vero	7	0,7	8	0,8	0,75	-0,1
	falso	3	0,3	1	0,1	0,2	0,2
	O	-	0	1	0,1	0,05	-0,1
4	vero	9	0,9	7	0,7	0,8	0,2
	falso	1	0,1	2	0,2	0,15	-0,1
	O	-	0	1	0,1	0,05	-0,1
5	vero	1	0,1	4	0,4	0,25	-0,3
	falso	9	0,9	5	0,5	0,7	0,4
	O	-	0	1	0,1	0,05	-0,1
Parte quarta							
1	a	-	0	4	0,4	0,2	-0,4
	b	-	0	-	0	0	0
	c	10	1	6	0,6	0,8	0,4
	O	-	0	-	0	0	0
2	a	-	0	1	0,1	0,05	-0,1
	b	10	1	8	0,8	0,9	0,2
	c	-	0	1	0,1	0,05	-0,1
	O	-	0	-	0	0	0
3	a	1	0,1	1	0,1	0,1	0
	b	8	0,8	8	0,8	0,8	0
	c	1	0,1	1	0,1	0,1	0
	O	-	0	-	0	0	0
4	a	6	0,6	4	0,4	0,5	0,2
	b	4	0,4	5	0,5	0,45	-0,1
	c	-	0	1	0,1	0,05	-0,1
	O	-	0	-	0	0	0
5	a	10	1	9	0,9	0,95	0,1
	b	-	0	1	0,1	0,05	-0,1
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
6	a	10	1	7	0,7	0,85	0,3
	b	-	0	-	0	0	0
	c	-	0	2	0,2	0,1	-0,2
	O	-	0	1	0,1	0,05	-0,1
7	a	1	0,1	7	0,7	0,4	-0,6
	b	8	0,8	1	0,1	0,45	0,7
	c	1	0,1	2	0,2	0,15	-0,1
	O	-	0	-	0	0	0
8	a	-	0	1	0,1	0,05	-0,1
	B	-	0	2	0,2	0,1	-0,2
	c	10	1	4	0,4	0,7	0,6
	O	-	0	3	0,3	0,15	-0,3
9	a	5	0,5	4	0,4	0,45	0,1
	b	2	0,2	2	0,2	0,2	0
	c	3	0,3	2	0,2	0,25	0,1
	O	-	0	2	0,2	0,1	-0,2
10	a	-	0	5	0,5	0,25	-0,5
	b	10	1	2	0,2	0,6	0,8
	c	-	0	1	0,1	0,05	-0,1
	O	-	0	2	0,2	0,1	-0,2

Item Tally Sheet - Prova di Reading n. 2

Tipo di prova: lettura		Gruppo: B		Totale candidati: 18		Data: 31 marzo 2009	
N. item	Risp	Candidati gruppo B _n :9	p	Candidati gruppo B _i :9	p	IF	ID
Parte prima							
1	a	2	0,23	1	0,12	0,17	0,11
	b	1	0,12	3	0,34	0,23	-0,22
	<u>c</u>	6	0,67	5	0,56	0,61	0,11
	O	-	0	-	0	0	0
2	a	-	0	2	0,23	0,11	-0,23
	<u>b</u>	8	0,89	3	0,34	0,61	0,55
	c	1	0,12	4	0,45	0,28	-0,33
	O	-	0	-	0	0	0
3	a	1	0,12	1	0,12	0,12	0
	b	2	0,23	7	0,78	0,5	-0,55
	<u>c</u>	6	0,67	1	0,12	0,39	0,55
	O	-	0	-	0	0	0
4	a	2	0,23	-	0	0,11	0,23
	<u>b</u>	7	0,78	7	0,78	0,78	0
	c	-	0	2	0,23	0,11	-0,23
	O	-	0	-	0	0	0
5	<u>a</u>	8	0,89	4	0,45	0,67	0,44
	b	-	0	3	0,34	0,17	-0,34
	c	1	0,12	2	0,23	0,17	-0,11
	O	-	0	-	0	0	0
Parte seconda							
1	a	-	0	3	0,34	0,17	-0,34
	b	-	0	-	0	0	0
	<u>c</u>	9	1	6	0,67	0,83	0,33
	O	-	0	-	0	0	0
2	a	-	0	-	0	0	0
	<u>b</u>	9	1	4	0,45	0,72	0,55
	c	-	0	5	0,56	0,28	-0,56
	O	-	0	-	0	0	0
3	a	2	0,23	2	0,23	0,23	0
	<u>b</u>	7	0,78	2	0,23	0,5	0,55
	c	-	0	5	0,56	0,28	-0,56
	O	-	0	-	0	0	0
4	<u>a</u>	9	1	3	0,34	0,67	0,66
	b	-	0	2	0,23	0,11	-0,23
	c	-	0	4	0,45	0,22	-0,45
	O	-	0	-	0	0	0
5	a	6	0,67	7	0,78	0,72	-0,11
	<u>b</u>	2	0,23	2	0,23	0,23	0
	c	1	0,12	-	0	0,06	0,12
	O	-	0	-	0	0	0
Parte terza							
1	vero	-	0	6	0,67	0,33	-0,67
	<u>falso</u>	9	1	2	0,23	0,61	0,77
	O	-	0	1	0,12	0,06	-0,12
2	vero	2	0,23	6	0,67	0,45	-0,44
	<u>falso</u>	7	0,78	2	0,23	0,5	0,55

	O	-	0	1	0,12	0,06	-0,12
3	<u>vero</u>	7	0,78	1	0,12	0,45	0,66
	falso	2	0,23	7	0,78	0,50	-0,55
	O	-	0	1	0,12	0,06	-0,12
4	<u>vero</u>	9	1	6	0,67	0,83	0,33
	falso	-	0	2	0,23	0,11	-0,23
	O	-	0	1	0,12	0,06	-0,12
5	<u>vero</u>	3	0,34	5	0,56	0,45	-0,22
	falso	6	0,67	3	0,34	0,5	0,33
	O	-	0	1	0,12	0,06	-0,12
Parte quarta							
1	<u>a</u>	3	0,34	5	0,56	0,45	-0,22
	b	6	0,67	3	0,34	0,5	0,33
	<u>c</u>	-	0	-	0	0	0
	O	-	0	1	0,12	0,06	-0,12
2	<u>a</u>	9	1	4	0,45	0,72	0,55
	b	-	0	1	0,12	0,06	-0,12
	c	-	0	3	0,34	0,17	-0,34
	O	-	0	1	0,12	0,06	-0,12
3	<u>a</u>	-	0	-	0	0	0
	b	-	0	1	0,12	0,06	-0,12
	c	9	1	7	0,78	0,89	0,22
	O	-	0	1	0,12	0,06	-0,12
4	<u>a</u>	5	0,56	2	0,23	0,39	0,33
	b	-	0	1	0,12	0,06	-0,12
	c	-	0	5	0,56	0,28	-0,56
	O	4	0,45	1	0,12	0,28	0,33
5	<u>a</u>	-	0	2	0,23	0,11	-0,23
	b	9	1	3	0,34	0,67	0,66
	c	-	0	3	0,34	0,17	-0,34
	O	-	0	1	0,12	0,06	-0,12
6	<u>a</u>	8	0,89	1	0,12	0,5	0,77
	b	-	0	5	0,56	0,28	-0,56
	c	1	0,12	2	0,23	0,17	-0,11
	O	-	0	1	0,12	0,06	-0,12
7	<u>a</u>	1	0,12	2	0,23	0,17	-0,11
	b	8	0,89	2	0,23	0,56	0,66
	c	-	0	3	0,34	0,17	-0,34
	O	-	0	2	0,23	0,11	-0,23
8	<u>a</u>	-	0	-	0	0	0
	b	9	1	3	0,34	0,68	0,66
	c	-	0	4	0,45	0,22	-0,45
	O	-	0	2	0,23	0,11	-0,23
9	<u>a</u>	3	0,34	3	0,34	0,34	0
	b	6	0,67	4	0,45	0,56	0,22
	c	-	0	-	0	0	0
	O	-	0	2	0,23	0,11	-0,23
10	<u>a</u>	-	0	3	0,34	0,17	-0,34
	b	3	0,34	3	0,34	0,34	0
	c	-	0	1	0,12	0,06	-0,12
	O	6	0,67	2	0,23	0,45	0,44

Item Tally Sheet - Prova di Reading n. 3

Tipo di prova: lettura		Gruppo: C		Totale candidati: 18		Data: 31 marzo 2009	
N. item	Risp	Candidati gruppo C _h :9	p	Candidati gruppo C _i :9	p	IF	ID
Parte prima							
1	a	-	0	5	0,56	0,28	-0,56
	<u>b</u>	9	1	3	0,34	0,67	0,66
	c	-	0	1	0,12	0,06	-0,12
	O	-	0	-	0	0	0
2	a	3	0,34	7	0,78	0,56	-0,44
	b	-	0	-	0	0	0
	<u>c</u>	6	0,67	2	0,23	0,45	0,44
	O	-	0	-	0	0	0
3	<u>a</u>	9	1	9	1	1	0
	b	-	0	-	0	0	0
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
4	a	-	0	-	0	0	0
	<u>b</u>	9	1	9	1	1	0
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
5	a	-	0	1	0,12	0,06	-0,12
	b	1	0,12	5	0,56	0,34	-0,44
	<u>c</u>	8	0,89	3	0,34	0,61	0,55
	O	-	0	-	0	0	0
Parte seconda							
1	a	-	0	-	0	0	0
	b	1	0,12	4	0,45	0,28	-0,33
	<u>c</u>	8	0,89	4	0,45	0,67	0,44
	O	-	0	1	0,12	0,06	-0,12
2	<u>a</u>	9	1	5	0,56	0,78	0,44
	b	-	0	1	0,12	0,06	-0,12
	c	-	0	2	0,23	0,11	-0,23
	O	-	0	1	0,12	0,06	-0,12
3	a	-	0	1	0,12	0,06	-0,12
	b	-	0	-	0	0	0
	<u>c</u>	9	1	6	0,67	0,83	0,33
	O	-	0	2	0,23	0,11	-0,23
4	a	-	0	-	0	0	0
	b	3	0,34	2	0,23	0,28	0,11
	<u>c</u>	6	0,67	3	0,34	0,5	0,33
	O	-	0	4	0,45	0,22	-0,45
5	<u>a</u>	3	0,34	1	0,12	0,23	0,22
	b	-	0	-	0	0	0
	c	6	0,67	5	0,56	0,61	0,11
	O	-	0	3	0,34	0,17	-0,34
Parte terza							
1	vero	8	0,89	7	0,78	0,83	0,11
	<u>falso</u>	1	0,12	2	0,23	0,17	-0,11
	O	-	0	-	0	0	0
2	vero	-	0	5	0,56	0,28	-0,56
	<u>falso</u>	9	1	4	0,45	0,72	0,55

	O	-	0	-	0	0	0
3	vero	-	0	1	0,12	0,06	-0,12
	falso	9	1	8	0,89	0,94	0,11
	O	-	0	-	0	0	0
4	vero	7	0,78	5	0,56	0,67	0,22
	falso	2	0,23	4	0,45	0,34	-0,22
	O	-	0	-	0	0	0
5	vero	8	0,89	9	1	0,94	-0,11
	falso	1	0,12	-	0	0,06	0,12
	O	-	0	-	0	0	0
Parte quarta							
1	a	-	0	-	0	0	0
	b	1	0,12	-	0	0,06	0,12
	c	8	0,89	8	0,89	0,89	0
	O	-	0	1	0,12	0,06	-0,12
2	a	1	0,12	7	0,78	0,45	-0,66
	b	-	0	-	0	0	0
	c	8	0,89	1	0,12	0,5	0,77
	O	-	0	1	0,12	0,06	-0,12
3	a	-	0	-	0	0	0
	b	9	1	8	0,89	0,94	0,11
	c	-	0	-	0	0	0
	O	-	0	1	0,12	0,06	-0,12
4	a	-	0	-	0	0	0
	b	5	0,56	7	0,78	0,67	-0,22
	c	4	0,45	1	0,12	0,28	0,33
	O	-	0	1	0,12	0,06	-0,12
5	a	9	1	6	0,67	0,83	0,33
	b	-	0	2	0,23	0,11	-0,23
	c	-	0	-	0	0	0
	O	-	0	1	0,12	0,06	-0,12
6	a	9	1	7	0,78	0,89	0,22
	b	-	0	-	0	0	0
	c	-	0	-	0	0	0
	O	-	0	2	0,23	0,11	-0,23
7	a	3	0,34	3	0,34	0,34	0
	b	-	0	-	0	0	0
	c	6	0,67	5	0,56	0,61	0,11
	O	-	0	1	0,12	0,06	-0,12
8	a	-	0	-	0	0	0
	b	-	0	2	0,23	0,11	-0,23
	c	9	1	6	0,67	0,83	0,33
	O	-	0	1	0,12	0,06	-0,12
9	a	-	0	5	0,56	0,28	-0,56
	b	9	1	2	0,23	0,61	0,77
	c	-	0	-	0	0	0
	O	-	0	2	0,23	0,11	-0,23
10	a	-	0	-	0	0	0
	b	9	1	6	0,67	0,83	0,33
	c	-	0	1	0,12	0,06	-0,12
	O	-	0	2	0,23	0,11	-0,23

Item Tally Sheet - Prova di Reading n. 4

Tipo di prova: lettura		Gruppo: D		Totale candidati: 18		Data: 31 marzo 2009	
N. item	Risp	Candidati gruppo D _h :9	p	Candidati gruppo D _i :9	p	IF	ID
Parte prima							
1	a	-	0	1	0,12	0,06	-0,12
	b	2	0,23	8	0,89	0,56	-0,66
	<u>c</u>	7	0,78	-	0	0,39	0,78
	O	-	0	-	0	0	0
2	a	2	0,23	4	0,45	0,34	-0,22
	<u>b</u>	7	0,78	5	0,56	0,67	0,22
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
3	<u>a</u>	9	1	3	0,34	0,67	0,66
	b	-	0	-	0	0	0
	c	-	0	6	0,67	0,33	-0,67
	O	-	0	-	0	0	0
4	a	1	0,12	-	0	0,06	0,12
	<u>b</u>	8	0,89	4	0,45	0,67	0,44
	c	-	0	5	0,56	0,28	-0,56
	O	-	0	-	0	0	0
5	a	-	0	1	0,12	0,06	-0,12
	b	-	0	-	0	0	0
	<u>c</u>	9	1	8	0,89	0,94	0,11
	O	-	0	-	0	0	0
Parte seconda							
1	<u>a</u>	9	1	6	0,67	0,83	0,33
	b	-	0	-	0	0	0
	c	-	0	3	0,34	0,17	-0,34
	O	-	0	-	0	0	0
2	a	-	0	-	0	0	0
	b	-	0	1	0,12	0,06	-0,12
	<u>c</u>	9	1	8	0,89	0,94	0,11
	O	-	0	-	0	0	0
3	a	-	0	-	0	0	0
	b	-	0	-	0	0	0
	<u>c</u>	9	1	9	1	1	0
	O	-	0	-	0	0	0
4	a	-	0	-	0	0	0
	<u>b</u>	8	0,89	9	1	0,94	-0,11
	c	1	0,12	-	0	0,06	0,12
	O	-	0	-	0	0	0
5	<u>a</u>	9	1	5	0,56	0,78	0,44
	b	-	0	2	0,23	0,11	-0,23
	c	-	0	2	0,23	0,11	-0,23
	O	-	0	-	0	0	0
Parte terza							
1	vero	1	0,12	7	0,78	0,45	-0,66
	<u>falso</u>	8	0,89	2	0,23	0,56	0,66
	O	-	0	-	0	0	0
2	<u>vero</u>	9	1	6	0,67	0,83	0,33
	falso	-	0	3	0,34	0,17	-0,34

	O	-	0	-	0	0	0
3	<u>vero</u>	9	1	8	0,89	0,94	0,11
	falso	-	0	1	0,12	0,06	-0,12
	O	-	0	-	0	0	0
4	<u>vero</u>	9	1	8	0,89	0,94	0,11
	falso	-	0	1	0,12	0,06	-0,12
	O	-	0	-	0	0	0
5	<u>vero</u>	-	0	2	0,23	0,11	-0,23
	falso	9	1	7	0,78	0,89	0,22
	O	-	0	-	0	0	0
Parte quarta							
1	a	-	0	1	0,12	0,06	-0,12
	<u>b</u>	7	0,78	6	0,67	0,72	0,11
	c	2	0,23	2	0,23	0,23	0
	O	-	0	-	0	0	0
2	<u>a</u>	9	1	2	0,23	0,61	0,77
	b	-	0	2	0,23	0,11	-0,23
	c	-	0	4	0,45	0,22	-0,45
	O	-	0	1	0,12	0,06	-0,12
3	a	7	0,78	2	0,23	0,5	0,55
	<u>b</u>	-	0	1	0,12	0,06	-0,12
	c	2	0,23	5	0,56	0,39	-0,33
	O	-	0	1	0,12	0,06	-0,12
4	a	-	0	1	0,12	0,06	-0,12
	b	-	0	2	0,23	0,11	-0,23
	<u>c</u>	9	1	5	0,56	0,78	0,44
	O	-	0	1	0,12	0,06	-0,12
5	a	-	0	3	0,34	0,17	-0,34
	<u>b</u>	9	1	5	0,56	0,78	0,44
	c	-	0	-	0	0	0
	O	-	0	1	0,12	0,06	-0,12
6	a	-	0	-	0	0	0
	<u>b</u>	9	1	8	0,89	0,94	0,11
	c	-	0	-	0	0	0
	O	-	0	1	0,12	0,06	-0,12
7	<u>a</u>	9	1	3	0,34	0,66	0,66
	b	-	0	-	0	0	0
	c	-	0	3	0,34	0,17	-0,34
	O	-	0	3	0,34	0,17	-0,34
8	a	1	0,12	2	0,23	0,17	-0,11
	<u>b</u>	8	0,89	2	0,23	0,56	0,66
	c	-	0	3	0,34	0,17	-0,34
	O	-	0	2	0,23	0,11	-0,23
9	a	-	0	1	0,12	0,06	-0,12
	<u>b</u>	9	1	3	0,34	0,67	0,66
	c	-	0	3	0,34	0,17	-0,34
	O	-	0	2	0,23	0,11	-0,23
10	a	-	0	3	0,34	0,17	-0,34
	b	-	0	1	0,12	0,06	-0,12
	<u>c</u>	9	1	3	0,34	0,67	0,66
	O	-	0	2	0,23	0,11	-0,23

Item Tally Sheet - Prova di Reading n. 5

Tipo di prova: lettura		Gruppo: E		Totale candidati: 10		Data: 29 giugno 2009	
N. item	Risp	Candidati gruppo E _n :5	p	Candidati gruppo E _i :5	p	IF	ID
Parte prima							
1	a	-	0	-	0	0	0
	b	-	0	3	0,6	0,3	-0,6
	<u>c</u>	5	1	2	0,4	0,7	0,6
	O	-	0	-	0	0	0
2	a	-	0	3	0,6	0,3	-0,6
	<u>b</u>	5	1	2	0,4	0,7	0,6
	c	-	0	-	0	0	0
	O	-	0	-	0	0	0
3	<u>a</u>	5	1	3	0,6	0,8	0,4
	b	-	0	1	0,2	0,1	-0,2
	c	-	0	1	0,2	0,1	-0,2
	O	-	0	-	0	0	0
4	<u>a</u>	3	0,6	3	0,6	0,6	0
	b	2	0,4	1	0,2	0,3	0,2
	c	-	0	1	0,2	0,1	-0,2
	O	-	0	-	0	0	0
5	a	-	0	2	0,4	0,2	-0,4
	b	-	0	1	0,2	0,1	-0,2
	<u>c</u>	5	1	2	0,4	0,7	0,6
	O	-	0	-	0	0	0
Parte seconda							
1	a	1	0,2	-	0	0,1	0,2
	b	4	0,8	5	1	0,9	-0,2
	<u>c</u>	-	0	-	0	0	0
	O	-	0	-	0	0	0
2	a	1	0,2	4	0,8	0,5	-0,6
	b	3	0,6	1	0,2	0,4	0,4
	<u>c</u>	1	0,2	-	0	0,1	0,2
	O	-	0	-	0	0	0
3	a	-	0	2	0,4	0,2	-0,4
	<u>b</u>	3	0,6	1	0,2	0,4	0,4
	c	2	0,4	2	0,4	0,4	0
	O	-	0	-	0	0	0
4	a	2	0,4	4	0,8	0,6	-0,4
	b	-	0	1	0,2	0,1	-0,2
	<u>c</u>	3	0,6	-	0	0,3	0,6
	O	-	0	-	0	0	0
5	a	-	0	2	0,4	0,2	-0,4
	<u>b</u>	3	0,6	1	0,2	0,4	0,4
	c	2	0,4	2	0,4	0,4	0
	O	-	0	-	0	0	0
Parte terza							
1	vero	-	0	3	0,6	0,3	-0,6
	<u>falso</u>	5	1	2	0,4	0,7	0,6
	O	-	0	-	0	0	0
2	<u>vero</u>	1	0,2	-	0	0,1	0,2
	falso	4	0,8	5	1	0,9	-0,2

	O	-	0	-	0	0	0
3	vero	-	0	3	0,6	0,3	-0,6
	<u>falso</u>	5	1	2	0,4	0,7	0,6
	O	-	0	-	0	0	0
4	vero	1	0,2	3	0,6	0,4	-0,4
	<u>falso</u>	4	0,8	2	0,4	0,6	0,4
	O	-	0	-	0	0	0
5	vero	2	0,4	3	0,6	0,5	-0,2
	<u>falso</u>	3	0,6	1	0,2	0,4	0,4
	O	-	0	1	0,2	0,1	-0,2
Parte quarta							
1	a	4	0,8	1	0,2	0,5	0,6
	b	-	0	3	0,6	0,3	-0,6
	<u>c</u>	1	0,2	-	0	0,1	0,2
	O	-	0	1	0,2	0,1	-0,2
2	<u>a</u>	5	1	3	0,6	0,8	0,4
	b	-	0	-	0	0	0
	c	-	0	-	0	0	0
	O	-	0	2	0,4	0,2	-0,4
3	a	-	0	1	0,2	0,1	-0,2
	<u>b</u>	5	1	3	0,6	0,8	0,4
	c	-	0	-	0	0	0
	O	-	0	1	0,2	0,1	-0,2
4	a	2	0,4	3	0,6	0,5	-0,2
	b	-	0	-	0	0	0
	<u>c</u>	3	0,6	-	0	0,3	0,6
	O	-	0	2	0,4	0,2	-0,4
5	<u>a</u>	5	1	-	0	0,5	1
	b	-	0	1	0	0,1	-0,2
	c	-	0	1	0,2	0,1	-0,2
	O	-	0	3	0,2	0,3	-0,6
6	a	-	0	-	0	0	0
	<u>b</u>	5	1	3	0,6	0,8	0,4
	c	-	0	-	0	0	0
	O	-	0	2	0,4	0,2	-0,4
7	a	-	0	2	0,4	0,2	-0,4
	b	-	0	-	0	0	0
	<u>c</u>	5	1	-	0	0,5	1
	O	-	0	3	0,6	0,3	-0,6
8	<u>a</u>	5	1	1	0,2	0,6	0,8
	b	-	0	-	0	0	0
	c	-	0	-	0	0	0
	O	-	0	4	0,8	0,4	-0,8
9	<u>a</u>	5	1	1	0,2	0,6	0,8
	b	-	0	-	0	0	0
	c	-	0	1	0,2	0,1	-0,2
	O	-	0	3	0,6	0,3	-0,6
10	a	2	0,4	1	0,2	0,3	0,2
	<u>b</u>	3	0,6	-	0	0,3	0,6
	c	-	0	1	0,2	0,1	-0,2
	O	-	0	3	0,6	0,3	-0,6

RIFERIMENTI BIBLIOGRAFICI

- ALDERSON, J.C. (2001), "The shape of things to come: will it be the normal distribution?", in Milanovic, M. e Weir, C.J., pp. 1-26.
- ALDERSON, J.C. (2000), *Assessing Reading*, Cambridge, Cambridge University Press.
- ALDERSON, J.C., BACHMAN, L.F., PERKINS, K., COHEN, A. (1991), "An exploratory study into the construct validity of a reading comprehension test: triangulation of data sources", in *Language Testing* 8 (1), pp. 41-66.
- ALDERSON, J.C., CLAPHAM, C., WALL, D. (1995), *Language Test Construction and Evaluation*, Cambridge, Cambridge University Press.
- ALDERSON, J.C. (1993), "Judgments in Language Testing", in Douglas, D. e Chapelle, C., pp. 46-57.
- ALDERSON, J.C. e NORTH, B. (1991), *Language Testing in the 1990s: the Communicative Legacy*, London, Macmillan.
- ALTE, *Materials for the Guidance of Test Item Writers*, <http://www.alte.org/downloads/index.php> (ultima consultazione: febbraio 2011).
- ALTE, *Codice professionale dell'ALTE*, <http://www.alte.org/downloads/index.php?docid=23> (ultima consultazione: febbraio 2011).
- ALTE, *Code of Practice*, <http://www.alte.org/downloads/index.php?doctypeid=6> (ultima consultazione: febbraio 2011).
- ALTE, *Can Do statements*, http://www.coe.int/t/dg4/linguistic/source/Framework_EN.pdf (ultima consultazione: febbraio 2011).

- ALTE (1998), *Multilingual Glossary of Language Testing Terms* (Studies in Language Testing 6), Cambridge, UCLES/Cambridge University Press.
- ALTE (1998), *Handbook of Language Examinations and Examination Systems*, Cambridge, Cambridge, Cambridge University Press.
- ALTE (1997), *User Guide for Examiners*, ristampato dal Consiglio D'Europa (2002), *Language Examining and Test Development*.
- AMERICAN PSYCHOLOGICAL ASSOCIATION (1985a), *Standards for Educational and Psychological Testing*, Washington, DC, American Psychological Association.
- AMERICAN PSYCHOLOGICAL ASSOCIATION (1985b), "Technical recommendations for psychological tests and diagnostic techniques", in *Psychological Bulletin*, 51 (2:2), pp. 1-38.
- AMERICAN PSYCHOLOGICAL ASSOCIATION (1974), *Standards for Educational and Psychological Testing*, Washington, DC, American Psychological Association.
- ANCKAR, J. (2006), *Guessing and elimination at a multiple-choice test of listening comprehension*, presentato alla terza conferenza annuale dell'EALTA, Cracovia.
- ARENA, L.A. (1990), *Language Proficiency, Defining, Teaching and Testing*, New York and London, Plenum Press.
- BACHMAN, L.F. (2005), *Statistical Analyses for Language Assessment Workbook*, Cambridge, Cambridge University Press.
- BACHMAN, L.F. e KUNNAN, A.J. (2004), *Statistical Analyses for Language Assessment and CD ROM*, Cambridge, Cambridge University Press.

- BACHMAN, L.F. e COHEN, A.D. (1998), *Interfaces between Second Language Acquisition and Language Testing Research*, Cambridge, Cambridge University Press.
- BACHMAN, L.F. e PALMER, A. (1996), *Language Testing in Practice*, Oxford, Oxford University Press.
- BACHMAN, L.F. (1990), *Fundamental Considerations in Language Testing*, Oxford, Oxford University Press.
- BAILEY, K.M. (1997), *Learning about Language Assessment: dilemmas, decisions and directions*, Boston, Heinle and Heinle.
- BAKER, R. (1997), "Classical test theory and item response theory in test analysis", in *An investigation of the Rasch model in its application to foreign language proficiency testing*, Special Report 2: Language Testing Update, Centre for Research in Language Education, Lancaster, Lancaster University.
- BALBONI, P.E. (2002), *Le sfide di Babele. Insegnare le lingue nelle società complesse*, Torino, UTET.
- BALBONI, P.E. (1999), *Dizionario di glottodidattica*, Perugia, Guerra-Soleil.
- BIFERALE, N., *Quale profilo e quali competenze per l'inglese del docente di scuola primaria - Un'esperienza di ricerca*, <http://puntoeduri.indire.it/lingue> (ultima consultazione: febbraio 2011).
- BOSISIO, C. (2007a), "Insegnare oggi una lingua "al tra": tra quotidianità didattica e realtà europea", in Chini, M., Desideri, P., Favilla, E., Pallotti, G., pp. 375-410.
- BOSISIO, C. (2007b), "Insegnare lingue in Italia: questioni di educazione

- plurilinguistica”, in Gomez-Pablos, B., Ollivier, C., *Aktuelle Tendenzen in der romanistischen Didaktik*, Amburgo, Kovac, pp.111-132.
- BOSISIO, C. (2005), *Dagli approcci tradizionali al Quadro comune europeo di riferimento. Riflessioni glottodidattiche e applicazioni per l'insegnante di italiano L2*, Milano, EDUCatt Università Cattolica.
- BOSISIO, C. (2005b), “Insegnare italiano L2 oggi: verso un percorso formativo ideale”, in Cambiaghi, B., Milani, C., Pontani, P. (a cura di), *Europa plurilingue. Comunicazione e didattica. Atti del Convegno Internazionale di Studi*, Milano, Università Cattolica del Sacro Cuore, Vita e Pensiero, pp. 203-227.
- BROWN, G., YULE, G. (1983), *Discourse Analysis*, Cambridge, Cambridge University Press.
- BROWN, J.D. e HUDSON, T. (2002), *Criterion-referenced Language Testing*, Cambridge, Cambridge University Press.
- BROWN, J.D. (2005), *Testing in Language Programs: a comprehensive guide to English language assessment*, New York, McGraw-Hill.
- BROWN, J.D. (1996), *Testing in Language Programs*, New Jersey, Prentice Hall.
- BUCK, G. (2001), *Assessing Listening*, Cambridge, Cambridge University Press.
- CAMPBELL, D.T. e FISKE, D.W. (1959), “Convergent and discriminant validation by the multitrait-multimethod matrix”, in *Psychological Bulletin*, 56 (2), pp. 81-105.
- CANALE, M. e SWAIN, M. (1980), “Theoretical bases of communicative approaches to second language teaching and testing”, in *Applied Linguistics* 1 (1), pp. 1-47.

- CANDLIN, C. (1986), "Explaining communicative competence limits of testability", in Stansfield, C., *Towards Communicative Competence Testing: proceedings of the second TOEFLS invitational conference*, Princeton, New Jersey, Educational Testing Service, pp. 38-57.
- CARITAS MIGRANTES, *Immigrazione. Dossier statistico 2009, XIX rapporto*, <http://www.dossierimmigrazione.it/schede/pres2009--scheda.pdf> (ultima consultazione: febbraio 2011).
- CARROLL, J.B. (1980), *Testing Communicative Performance*, Oxford, Pergamon Press.
- CARROLL, J.B. (1978), *Guidelines for the Development of Communicative Tests*, London, Royal Society of Arts.
- CARROLL, J.B. (1968), "The psychology of language testing" in Davies, A. (a cura di), *Language Testing Symposium. A Psycholinguistic Perspective*, London, Oxford University Press, pp. 46-69.
- CARROLL, J.B. (1961), "Fundamental considerations in testing English language proficiency of foreign students", in *Testing the English Proficiency of Foreign Students*, Washington, DC, Center for Applied Linguistics, pp. 30-40, ristampato in Allen, H.B. e Campbell, R.N. (1972), *Teaching English as a Second Language: a book of readings*, New York, McGraw-Hill, pp. 313-321.
- CHALHOUB-DEVILLE, M., CHAPELLE, C., DUFF, P. (2006), *Inference and Generalizability in Applied Linguistics. Multiple Perspectives*, Amsterdam, John Benjamins.
- CHALHOUB-DEVILLE, M. (1999), *Issues in Computer-Adaptive Testing of Reading Proficiency* (Studies in Language Testing 10), Cambridge,

Cambridge University Press.

CHAPELLE, C. e DOUGLAS, D. (2006), *Assessing Language through Computer Technology*, Cambridge, Cambridge University Press.

CHENG, L. (2007), *Language Testing Reconsidered*, Ottawa, On, University of Ottawa Press.

CHENG, L. (2003), *Washback in Language Testing: research contexts and methods*, New Jersey, Routledge.

CHINI, M., DESIDERI, P., FAVILLA, M.E., PALLOTTI, G. (a cura di) (2007), *Imparare una lingua: recenti sviluppi teorici e proposte applicative. Atti del VI Congresso Internazionale dell'Associazione Italiana di Linguistica Applicata*, Perugia, Guerra Edizioni.

CLAPHAM, C. e CORSON, D. (a cura di) (1997), *Language Testing and Assessment, Encyclopaedia of Language and Education*, Volume 7, Dordrecht, Kluwer Academic Publishers.

COMMISSIONE EUROPEA, *L'indicatore europeo di competenza linguistica*, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0356:FIN:IT:PDF> (ultima consultazione: febbraio 2011).

COMMISSIONE EUROPEA, *Libro bianco - Istruzione e formazione - Insegnare e apprendere - Verso la società conoscitiva*, http://europa.eu/documents/comm/white_papers/pdf/com95_590_en.pdf (ultima consultazione: febbraio 2011).

COMMISSIONE EUROPEA, *Libro verde - Istruzione su Formazione e Ricerca - Gli ostacoli alla mobilità transnazionale*, http://aei.pitt.edu/1226/01/education_mobility_obstacles_gp_COM_96_462.pdf (ultima consultazione: febbraio 2011).

COMMISSIONE EUROPEA, *Libro verde - Migrazione e mobilità: le sfide e le opportunità per i sistemi d'istruzione europei*, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2008:0423:FIN:IT:PDF> (ultima consultazione: febbraio 2011).

COMMISSIONE EUROPEA, *Libro verde - Promuovere la mobilità dei giovani per l'apprendimento*, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2009:0329:FIN:IT:PDF> (ultima consultazione: febbraio 2011).

COMMISSIONE EUROPEA, *Promuovere l'apprendimento delle lingue e la diversità linguistica: Piano d'azione 2004-2006*, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2003:0449:FIN:IT:PDF> (ultima consultazione: febbraio 2011).

COMMISSIONE EUROPEA, *Un nuovo quadro strategico per il multilinguismo*, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2005:0596:FIN:IT:PDF> (ultima consultazione: febbraio 2011).

COMUNITÀ EUROPEA, *Trattato che istituisce la Comunità Europea*, http://eur-lex.europa.eu/it/treaties/dat/12002E/pdf/12002E_IT.pdf (ultima consultazione: febbraio 2011).

CONSIGLIO DELL'UNIONE EUROPEA, *Quadro strategico per la cooperazione europea nel settore dell'istruzione e della formazione*, <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2009:119:0002:01:it:HTML> (ultima consultazione: febbraio 2011).

CONSIGLIO D'EUROPA, *Raccomandazioni e Risoluzioni*, http://www.coe.int/t/i/cplre/testi_in_italiano (ultima consultazione: febbraio 2011).

CONSIGLIO D'EUROPA, *La Convenzione culturale europea*, <http://www.admin.ch/ch/i/rs/i4/0.440.1.it.pdf> (ultima consultazione: febbraio

2011).

CONSIGLIO D'EUROPA, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, http://www.coe.int/t/dg4/linguistic/cadre_en.asp (ultima consultazione: febbraio 2011).

CONSIGLIO D'EUROPA (2001), *Quadro comune europeo di riferimento per le lingue: apprendimento insegnamento valutazione*, Oxford, La Nuova Italia.

CRONBACH, L.J. (1980), "Validity on parole: how can we go straight?", *New Directions for Testing and Measurement: measuring achievement over a decade. Proceedings of the 1979 ETS Invitational Conference*, 5, San Francisco, Jossey-Bass, pp. 99-108.

CRONBACH, L.J. (1971), "Test Validation", in Thorndike, R.L., pp. 443-507.

CRONBACH, L.J. e MEEHL, P.E. (1955), "Construct validity in psychological tests", in *Psychological Bulletin*, 52, pp. 281-302.

CUMMING, A. e BERWICK, R. (a cura di) (1996), *Validation in Language Testing*, Clevedon, Multilingual Matters.

CUMMINS, J. (1983), "Language proficiency and academic achievement", in Oller, J.W., pp. 108-130.

CUMMINS, J. (1979), "Linguistic interdependence and the educational development of bilingual children", in *Review of Educational Research* 49 (2), pp. 222-251.

DAVIDSON, F. (2000), "The language tester's statistical toolbox", in *System* 28 (4), pp. 605-617.

DAVIDSON, F. (1996), *Principles of Statistical Data Handling*, Thousand Oaks, CA, Sage Publications, Inc.

- DAVIDSON, F. e HENNING, G. (1985), "A self-rating scale of English difficulty: Rasch scalar analysis of items and rating categories", in *Language Testing* 2 (2), pp. 164-179.
- DAVIES, A. (2008), *Assessing Academic English: testing English proficiency 1950-1989 - the IELTS solution* (Studies in Language Testing 23), Cambridge, Cambridge University Press.
- DAVIES, A. (1990), *Principles of Language Testing*, Oxford, Blackwell.
- DAVIES, A., BROWN, A., ELDER, C., HILL, K., LUMLEY, T., MCNAMARA, T. (1999), *Dictionary of Language Testing* (Studies in Language Testing 7), Cambridge, Cambridge University Press.
- DE MAURO, T. (2000), *Dizionario di italiano*, Milano, Paravia Bruno Mondadori.
- DOUGLAS, D. (2010), *Understanding Language Testing*, London, Hodder Education.
- DOUGLAS, D. (2000), *Assessing Language for Specific Purposes*, Cambridge, Cambridge University Press.
- DOUGLAS, D. e CHAPELLE, C. (a cura di) (1993), *A New Decade of Language Testing Research: selected papers from the 1990 Language Testing Research Colloquium*, Alexandria, VA, TESOL Publications.
- FIELD, J.C. (2008), *Listening in the Second Language Classroom*, Cambridge, Cambridge University Press.
- FIELD, J.C. (2005), *Language and the Mind*, New York, Routledge.
- FULCHER, G. (2010), *Practical Language Testing*, London, Hodder Education.
- FULCHER, G. (1999), "Assessment in English for Academic Purposes: putting content validity in its place", in *Applied Linguistics* 20 (2), pp. 221-236.

- FULCHER, G. e DAVIDSON, F. (2007), *Language Testing and Assessment: an advanced resource book*, New York, Routledge.
- GERANPAYEH, A. (2008), "Using DIF to explore item difficulty in CAE Listening", in *Research Notes* 32, pp. 16-23.
- GERANPAYEH, A. e TAYLOR, L. (2008), "Examining Listening: developments and issues in assessing second language listening", in *Research Notes* 32, pp. 2-5.
- GERANPAYEH, A. (2001), "CB BULATS: examining the reliability of a computer based test using test-retest method", in *Research Notes* 5, pp. 14-16.
- GIPPS, C. (1994), *Beyond Testing. Towards a Theory of Educational Assessment*, London, Falmer Press.
- GOTTI, M., *Proposta per la formazione linguistica dei docenti in servizio nelle scuole dell'infanzia, primaria e secondaria di primo grado*, allegato al prot. n. 1446, http://archivio.pubblica.istruzione.it/normativa/2005/prot1446_05.shtml (ultima consultazione: febbraio 2011).
- HACKETT, E. (2008), "Adapting listening tests for on-screen use", in *Research Notes* 32, pp. 23-25.
- HALLIDAY, M.A.K. (1976), "The form of a functional grammar", in Kress, G., Halliday: *System and Function in Language*, London, Oxford University Press, pp. 7-25.
- HAWKEY, R. (2006), *Impact Theory and Practice: Studies of the IELTS Test and Progetto Lingue 2000* (Studies in Language Testing 24), Cambridge, Cambridge University Press.
- HAWKEY, R. (2004), *A Modular Approach to Testing English Language Skills:*

the development of the certificates in English language skills (CELS) examinations, Cambridge, Cambridge University Press.

HEATON, J.B. (1975), *Writing English Language Tests*, London and New York, Longman.

HENNING, G. (1987), *A Guide to Language Testing: development, evaluation, research*, Rowley, Mass, Newbury House.

HUGHES, A. (2003), *Testing for Language Teachers*, Cambridge, Cambridge University Press.

HYMES, D. (1972), "On communicative competence", in Pride, J. e Holmes, J. (a cura di), *Sociolinguistics*, Harmondsworth, Penguin, pp. 269-293.

INDIRE, *Formazione di Competenze Linguistico-Comunicative della Lingua Inglese dei Docenti della Scuola Primaria*, allegato al prot. n. 1446, http://archivio.pubblica.istruzione.it/normativa/2005/prot1446_05.shtml (ultima consultazione: febbraio 2011).

INDIRE e AICLU, *Protocollo di intesa tra INDIRE e AICLU*, http://archivio.pubblica.istruzione.it/normativa/2006/allegati/prot233_06.pdf (ultima consultazione: febbraio 2011).

JONES, N. (2002), "Relating the ALTE Framework to the Common European Framework of Reference", in CONSIGLIO D'EUROPA, *Case Studies on the use of the Common European Framework of Reference*, pp. 167-183.

JONES, N. e HIRTZEL, M. (2001), "Appendix D: The ALTE Can Do Statements", in Consiglio D'Europa, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Cambridge, Cambridge University Press, pp. 244-257.

- JONES, R.L. e SPOLSKY, B. (1975), *Testing language proficiency*, Arlington, Va, Center for Applied Linguistics.
- KELLY, M. e GRENFELL, M. (2004), *European Profile for Language Teacher Education – A Frame of Reference*, University of Southampton, Southampton, <http://www.lang.soton.ac.uk/profile/report/MainReport.pdf> (ultima consultazione: febbraio 2011).
- KERLINGER, F. (1986), *Foundations of Behavioral Research*, New York, Holt-Rinehart and Winston.
- KHALIFA, H. e WEIR, C.J. (2009), *Examining Reading: research and practice in assessing second language reading*, (Studies in Language Testing 29), Cambridge, Cambridge University Press.
- KRAMSCH, C. (1986), "From language proficiency to interactional competence", in *The Modern Language Journal*, 70 (4), pp. 366-372.
- KUNNAN, A.J. (1998), *Validation in Language Assessment*, Mahwah, New Jersey, Lawrence Erlbaum Associates.
- LADO, R. (1961), *Language Testing: the construction and use of foreign language tests*, London, Longman.
- LADO, R. (1957), *Linguistics across Cultures*, Ann Arbor, University of Michigan Press.
- LANGÉ, G. (2005), "Le lingue nella scuola europea: normative vigenti", in Cambiaghi, B., Milani, C., Pontani, P. (a cura di), *Europa plurilingue. Comunicazione e didattica. Atti del Convegno Internazionale di Studi*, Milano, Università Cattolica del Sacro Cuore, Vita e Pensiero, pp. 77-96.
- LARSEN-FREEMAN, D. e LONG, M.H. (1991), *An Introduction to Second*

Language Acquisition Research, London, Longman.

LAZARATON, A. (2005), "Quantitative Research Methods", in Hinkel, E., *Handbook of Research in Second Language Teaching and Learning*, New Jersey, Lawrence Erlbaum Associates, pp. 209-224.

LAZARATON, A. (1995), "Qualitative research in applied linguistics: a progress report", in *TESOL Quarterly* 29, pp. 455-472.

LITTLE, D. e PERCLOVÁ, R. (2001), *ELP: Guide for Teachers and Teacher Trainers*, http://www.coe.int/T/DG4/Portfolio/documents/ELPguide_teacherstrainers.pdf (ultima consultazione: febbraio 2011), Strasburgo, Consiglio d'Europa.

LYNCH, B.K. (2001), "Rethinking assessment from a critical perspective" in *Language Testing* 18 (4), pp. 351-372.

MACKIEWICZ, W. (2001), "Higher education and language policy in the European Union", in Milanovic, M. e Weir, C.J., pp. 173-186.

MARTYNIUK, W. (a cura di) (2010), *Aligning Tests with the CEFR: reflections on using the Council of Europe's draft manual* (Studies in Language Testing 33), Cambridge, Cambridge University Press.

MCNAMARA, T. e ROEVER, C. (2006), *Language Testing: the Social Dimension*, Oxford, Blackwell.

MCNAMARA, T. (2000), *Language Testing*, Oxford, Oxford University Press.

MCNAMARA, T. (1996), *Measuring Second Language Performance*, London and New York, Longman.

MESSICK, S. (1999), *Assessment in Higher Education: Issues of Access, Quality, Student Development, and Public Policy*, Mahwah, New Jersey,

Lawrence Erlbaum Associates.

MESSICK, S. (1989), "Validity", in Linn, R.L. (a cura di), *Educational Measurement*, New York, Macmillan, pp. 13-103.

MESSICK, S. e DOUGLAS, J. (1967), *Problems in Human Assessment*, New York, McGraw-Hill.

MILANOVIC, M. e WEIR, C.J. (a cura di) (2004), *European Language Testing in a Global Context, Proceedings of the ALTE Barcelona Conference July 2001* (Studies in Language Testing 18), Cambridge, Cambridge University Press.

MINISTERO DELL'ISTRUZIONE, DELL'UNIVERSITÀ E DELLA RICERCA, *Linee di orientamento per la formazione in servizio dei docenti di lingua inglese delle scuole dell'infanzia e primaria*, allegato al prot. n. 1446, http://archivio.pubblica.istruzione.it/normativa/2005/prot1446_05.shtml (ultima consultazione: febbraio 2011).

MINISTERO DELL'ISTRUZIONE, DELL'UNIVERSITÀ E DELLA RICERCA, *Piano Nazionale di Formazione per lo Sviluppo delle Competenze linguistico-comunicative e metodologico-didattiche in Lingua Inglese degli insegnanti di Scuola Primaria*, prot. n. 1446 del 19 luglio 2005, http://archivio.pubblica.istruzione.it/normativa/2005/prot1446_05.shtml (ultima consultazione: febbraio 2011).

MINISTERO DELL'ISTRUZIONE, DELL'UNIVERSITÀ E DELLA RICERCA, *Profilo professionale del docente di lingua inglese nelle scuole dell'infanzia e primaria*, allegato al prot. n. 1446, http://archivio.pubblica.istruzione.it/normativa/2005/prot1446_05.shtml (ultima consultazione: febbraio 2011).

- MISLEVY, R.J., STEINBERG, L.S., ALMOND, R.G. (2003), "On the Structure of Educational Assessments", in *Measurement: Interdisciplinary Research and Perspective* 1 (1), pp. 3-62.
- MORROW, K. (a cura di) (2004), *Insights from the Common European Framework*, Oxford, Oxford University Press.
- MOSIER, C.I. (1947), "A critical examination of the concepts of face validity", in *Educational and Psychological Measurement* 7 (2), pp. 191-205.
- MOSS, P. (1994), "Can There Be Validity Without Reliability?", in *Educational Researcher* 23 (2), pp. 5-12.
- O'LOUGHLIN, K. (2001), *The Equivalence of Direct and Semi-direct Speaking Tests* (Studies in Language Testing 13), Cambridge, Cambridge University Press.
- OLLER, J.W. Jr. (1983), *Issues in Language Testing Research*, Rowley, Mass. Newbury House.
- OLLER, J.W. Jr. (1972), "Cloze Tests of Second Language Proficiency and What They Measure", in *Language Learning* 23 (1), pp. 105-118.
- PARLAMENTO EUROPEO, CONSIGLIO E COMMISSIONE, *Carta dei diritti fondamentali dell'Unione Europea*, http://www.europarl.europa.eu/charter/pdf/text_it.pdf (ultima consultazione: febbraio 2011).
- PARLAMENTO EUROPEO, *Multilinguismo: una risorsa per l'Europa e un impegno comune*, <http://www.europarl.europa.eu/sides/getDoc.do?pubRef=-//EP//TEXT+TA+P6-TA-2009-0162+0+DOC+XML+V0//it> (ultima consultazione: febbraio 2011).
- PARLAMENTO ITALIANO, *Definizione delle norme generali relative alla scuola*

dell'infanzia e al primo ciclo dell'istruzione, Decreto Legislativo n. 59 del 19 febbraio 2004, <http://www.parlamento.it/parlam/leggi/deleghe/04059dl.htm> (ultima consultazione: febbraio 2011).

PARLAMENTO ITALIANO, *Delega al Governo per la definizione delle norme generali sull'istruzione e dei livelli essenziali delle prestazioni in materia di istruzione e formazione professionale*, Legge n. 53 del 28 marzo 2003, <http://www.parlamento.it/elenchileggi/43858/75871/elencoleggi.htm> (ultima consultazione: febbraio 2011).

PARLAMENTO ITALIANO, *Disposizioni per la formazione del bilancio annuale e pluriennale dello Stato (Finanziaria 2005)*, Legge n. 311 del 30 dicembre 2005, <http://www.parlamento.it/elenchileggi/43858/75870/elencoleggi.htm> (ultima consultazione: febbraio 2011).

PARLAMENTO ITALIANO, *Disposizioni per la formazione del bilancio annuale e pluriennale dello Stato (Finanziaria 2007)*, Legge n. 296 del 27 dicembre 2006, <http://www.parlamento.it/elenchileggi/43858/44080/elencoleggi.htm> (ultima consultazione: febbraio 2011).

PITTAU, F. (2009), *L'immigrazione nel XIX rapporto Caritas/Migrantes*, <http://www.dossierimmigrazione.it/schede/pres2009--pittau.pdf> (ultima consultazione: febbraio 2011).

PORCELLI, G. (1992), *Educazione linguistica e valutazione*, Padova, Liviana UTET.

PORCELLI, G. (1975), *Il Language Testing. Problemi e tecniche*, Bergamo, Minerva Italica.

RASCH, G. (1960), *Probabilistic models for some intelligence and attainment tests*, Chicago, University of Chicago Press.

- EBEL, R.L. (1961), "Must all tests be valid?" in *American Psychologist*, 16 (10), pp. 640-647.
- ROD, E. (2008), *The Study of Second Language Acquisition*, Oxford, Oxford University Press.
- ROD, E. (1997), *Second Language Acquisition*, Oxford, Oxford University Press.
- SAVIGNON, S.J. (1983), *Communicative Competence: Theory and Classroom Practice*, Reading, Mass, Addison-Wesley.
- SAVILLE, N. (2004), *The ESOL Test Development and Validation Strategy*, Cambridge, Cambridge University Press.
- SAVILLE, N. (2000), "Investigating the impact of international language examinations", in *Research Notes* 2, pp. 4-7.
- SCHNEIDER G., LENZ P., 2001, *ELP: Guide for Developers of an European Language Portfolio*, Strasburgo, Consiglio d'Europa, http://www.coe.int/T/DG4/Portfolio/documents_intro/Eguide.pdf (ultima consultazione: febbraio 2011).
- SHAW, S.D. (2005), *Item versus whole script marking in an ESM environment: a Cambridge Esol Perspective*, Cambridge, Cambridge University Press.
- SHOHAMY, E. (2000), "The relationship between language testing and second language acquisition", in *System* 28 (4), pp. 541-553.
- SPOLSKY, B. (1995), *Measured Words*, Oxford, Oxford University Press.
- SPOLSKY, B. (1986), "A multiple choice for language testers", in *Language Testing* 3 (2), pp. 147-158.
- TAYLOR, L. e FALVEY, P. (2007), *IELTS Collected Papers: research in speaking and writing assessment* (Studies in Language Testing 19), Cambridge,

Cambridge University Press.

TAYLOR, L. e JONES, N. (2006), "Cambridge ESOL exams and the Common European Framework of Reference", in *Research Notes* 24, pp. 2-5.

TAYLOR, L. e WEIR, C.J. (a cura di) (2008), *Multilingualism and Assessment: achieving transparency, assuring quality, sustaining diversity - Proceedings of the ALTE Berlin Conference, May 2005* (Studies in Language Testing 27), Cambridge, Cambridge University Press.

TAYLOR, L. (2004a), "Issues of test comparability", in *Research Notes* 15, pp. 2-5.

TAYLOR, L. (2004b), IELTS, "Cambridge ESOL examinations and the Common European Framework", in *Research Notes* 18, pp. 2-3.

TAYLOR, L. (1999), "Constituency matters: responsibilities and relationships in our test community", paper presented at the *Language Testing Forum*, Edinburgh.

TEASDALE, A. e LEUNG, C. (2000), "Teacher assessment and psychometric theory: a case of paradigm crossing?", in *Language Testing* 17 (2), pp. 163-184.

THORNDIKE, R.L. (1971), *Educational Measurement*, Washington, DC, American Council on Education.

TSAGARI, D. (2009), *The Complexity of Test Washback: an Empirical Study*, Frankfurt am Main, Peter Lang.

UFFICIO SCOLASTICO REGIONALE PER LA SICILIA, *Progetto Regionale per la Sicilia PFLISP*, http://62.77.63.181/ISN_Istruzionesicilia_it/pagina.aspx?pageid=a48b0bcd-6404-486c-89e6-d85d50d38426 (ultima consultazione:

febbraio 2011).

UNIVERSITY OF CAMBRIDGE ESOL EXAMINATIONS (2005), *PET Handbook for Teachers*, Cambridge, Cambridge University Press.

URQUHART, A.H. e ALDERSON, J.C. (1984), *Reading in a Foreign Language*, London, Longman.

URQUHART, A.H. e WEIR, C.J. (1998), *Reading in a Second Language: process, product and practice*. New York, Longman.

VALETTE, R.M. (1977), *Modern Language Testing*, New York, Harcourt Brace.

VAN DIJK, T.A. (1977), *Text and Context: Explorations in the Semantic and Pragmatics of Discourse*, London, Longman.

VISALBERGHI, A. e CORDA COSTA, M. (a cura di) (1995), *Misurare e valutare le competenze linguistiche*, Firenze, La Nuova Italia.

VISALBERGHI, A. (1955), *Misurazione e valutazione nel processo educativo*, Milano, Comunità (disponibile online: <http://diped.uniroma3.it/component/content/303?task=view>).

WEIR, C.J. (2005a), *Language Testing and Validation: an Evidence-Based Approach*, Basingstoke, Palgrave Macmillan.

WEIR, C.J. (2005b), "Limitation of the Common European Framework for developing comparable examinations and tests", in *Language Testing* 22 (3), pp. 281-300.

WEIR, C.J. (1993), *Understanding and Developing Language Tests*, New York and London, Prentice Hall.

WEIR, C.J. (1990), *Communicative Language Testing*, New York and London, Prentice Hall.

WEST, R. e TSAGARI, C. (2004), *Testing and Assessment in Language Learning*, Patras, Hellenic Open University.

WILSON, M. (2005), *Constructing measures: an item response modelling approach*, New Jersey, Lawrence Erlbaum Associates.

PERCORSI SITOGRAFICI

AICLU, Associazione Italiana Centri Linguistici Universitari, <http://www.aiclu.it>.

ALTE, *Association of Language Testers in Europe*, <http://www.alte.org>.

ANACLETO, <http://147.163.11.3/anacleto2008>.

CAMBRIDGE UNIVERSITY ESOL EXAMINATIONS, <http://www.cambridgeesol.org>.

CAMBRIDGE UNIVERSITY ESOL EXAMINATIONS, *Research Notes*,
http://www.cambridgeesol.org/rs_notes/.

CERCLU, Il progetto dell'AICLU per la Certificazione delle Competenze Linguistiche nelle Università Italiane, <http://www.aiclu.it/cerclu/nuovo/default.htm>.

CLAUSER, B.E., *Journal of Educational Measurement*, <http://www.wiley.com/bw/journal.asp?ref=0022-0655>.

CONSIGLIO D'EUROPA, <http://www.coe.int>.

DIALANG, <http://www.lancs.ac.uk/researchenterprise/dialang/about>.

EALTA, *European Association for Language Testing and Assessment*,
<http://www.ealta.eu.org>.

FULCHER, G., *Language Testing Resources Website*, <http://language-testing.info>.

ILTA, *International Language Testing Association*, <http://www.iltaonline.com>.

INSTITUTE FOR OBJECTIVE MEASUREMENT, <http://www.rasch.org/>.

OXFORD UNIVERSITY PRESS, UCLES (University of Cambridge Local Examinations Syndicate), *Quick Placement Test*, <http://www.britishcouncil.org/lithuania-education-uk-exams-in-lithuania-english-exams-placement.htm> (ultima consultazione: febbraio 2011) oppure <http://www.oxfordenglishtesting.com/Default MR.aspx?id=3034&menuId=1#> (ultima consultazione: febbraio 2011).

Portfolio Linguistico Europeo, <http://eelp.gap.it>.

Portfolio Linguistico Europeo elettronico EAQUALS-ALTE, <http://www.eelp.org/eportfolio/index.html>.

SPSS, <http://www.spss.it>.