

AVALIAÇÃO LINGUÍSTICA

GLOSSÁRIO ITALIANO / PORTUGUÊS

VALUTAZIONE LINGUISTICA

GLOSSARIO ITALIANO-PORTOGHESE

PAOLO TORRESAN



Pedro & João
editores

AVALIAÇÃO LINGUÍSTICA

GLOSSÁRIO ITALIANO/PORTUGUÊS

VALUTAZIONE LINGUISTICA

GLOSSARIO ITALIANO/PORTOGHESE

Paolo Torresan



Pedro & João
editores

AVALIAÇÃO LINGUÍSTICA

GLOSSÁRIO ITALIANO/PORTUGUÊS

VALUTAZIONE LINGUISTICA

GLOSSARIO ITALIANO/PORTOGHESE

Paolo Torresan

Copyright © Paolo Torresan

Todos os direitos garantidos. Qualquer parte desta obra pode ser reproduzida, transmitida ou arquivada desde que levados em conta os direitos do autor.

Paolo Torresan

Avaliação linguística: glossário italiano/português. Valutazione linguistica: glossario italiano/portoghese. São Carlos: Pedro & João Editores, 2022. 274p. 16 x 23 cm.

ISBN: 978-65-5869-739-8 [Digital]

1. Avaliação linguística. 2. Glossário italiano/português. 3. Habilidades linguísticas. 4. Metodologia de avaliação. I. Título.

CDD – 410

Capa: Petricor Design

Arte da folha de rosto: <https://www.iltaonline.com/>

Ficha Catalográfica: Hélio Márcio Pajeú – CRB - 8-8828

Tradutora: Carolina Giampietro

Contato com o autor: Paolo Torresan: piroclastico@gmail.com

Contato com a tradutora: carolinagiampietro@gmail.com

Diagramação: Diany Akiko Lee

Editores: Pedro Amaro de Moura Brito & João Rodrigo de Moura Brito

Conselho Científico da Pedro & João Editores:

Augusto Ponzio (Bari/Itália); João Wanderley Geraldi (Unicamp/ Brasil); Hélio Márcio Pajeú (UFPE/Brasil); Maria Isabel de Moura (UFSCar/Brasil); Maria da Piedade Resende da Costa (UFSCar/Brasil); Valdemir Miotello (UFSCar/Brasil); Ana Cláudia Bortolozzi (UNESP/Bauru/Brasil); Mariangela Lima de Almeida (UFES/Brasil); José Kuiava (UNIOESTE/Brasil); Marisol Barenco de Mello (UFF/Brasil); Camila Caracelli Schema (UFFS/Brasil); Luis Fernando Soares Zuin (USP/Brasil).



Pedro & João Editores

www.pedroejoaoeditores.com.br

13568-878 – São Carlos – SP

2022

תורה רבה

Apoio à publicação: Departamento de Letras Estrangeiras Modernas (GLE), UFF.

Le traduzioni dall'inglese sono dell'autore.
As traduções do inglês são feitas pelo autor.

Un ringraziamento a Carolina Giampietro, che ha tradotto il libro.
Um agradecimento a Carolina Giampietro, que traduziu o livro.

Indice

Índice

Prefazione dell'autore	29
Prefácio do autor	30
Introduzione (Guido Benvenuto)	31
Introdução (Guido Benvenuto)	35

PARTE I. COORDINATE FONDAMENTALI

PARTE I. COORDENADAS FUNDAMENTAIS

Cosa significa valutare?	41
O que significa avaliar?	42
Contesti valutativi	43
Contextos de avaliação	44
Tipi di prove	45
Tipos de prova	46
Effetti della valutazione	47
Efeitos da avaliação	48
Didattica e ricerca	49
Ensino e pesquisa	51
Valutazione linguistica: conoscenze vs abilità	53
Avaliação linguística: conhecimentos vs. habilidades	55

Valutare la comprensione	57
<i>Avaliar a compreensão</i>	62
Valutare la produzione	67
<i>Avaliar a produção</i>	72
Valutare la mediazione	76
<i>Avaliar a mediação</i>	78
Conclusioni	80
<i>Conclusões</i>	82

PARTE II. GLOSSARIO

PARTE II. GLOSSÁRIO

Introduzione	87
<i>Introdução</i>	89

A. CONCETTI GENERALI

A. CONCEITOS GERAIS

Abilità linguistica	91
<i>Habilidade linguística</i>	92
Argomento di validazione	93
<i>Argumento de validação</i>	93
Attività	94
<i>Atividade</i>	94
<i>Audit</i>	94
<i>Auditoria</i>	94

Ciclo di produzione	94
Ciclo de produção	95
Collegialità	95
Colegialidade	96
Compromesso	96
Compromisso	96
Conoscenza linguistica	97
Conhecimento linguístico	97
Competenza comunicativa	97
Competência comunicativa	97
Competenza specifica	98
Competência específica	98
Componente	98
Componente	98
Costrutto	99
Constructo	100
Dominio	101
Domínio	101
Estrapolazione	101
Extrapolação	101
<i>Feedback</i>	102
<i>Feedback</i>	102
<i>Performance</i>	102
<i>Performance</i>	103

Prova	103
Prova	103
Sillabo	103
Componentes curriculares	103
Specifiche	104
Especificações	104
<i>Standard</i>	104
Padrão	104
Strategia di apprendimento linguistico	104
Estratégia de aprendizagem linguística	104
Strategia metacognitiva	105
Estratégia metacognitiva	105
Strategia regolativa	105
Estratégia reguladora	105
Strategia di risoluzione di un test	106
Estratégia de resolução de um teste	106
Versione	106
Versão	106

B. ELEMENTI

B. ELEMENTOS

Batteria (di <i>item</i>)	107
Banco (de itens)	107
Chiave	107
Gabarito	107

Descrittore	107
Descriptor	107
Distrattore	107
Distrator	107
Griglia di analisi di un testo	108
Tabela de análise de um texto	108
Griglia di valutazione della complessità di un testo	108
Tabela de avaliação da complexidade de um texto	108
Griglia di valutazione della <i>performance</i>	108
Tabela de avaliação da <i>performance</i>	109
Informazione <i>target</i>	110
Informação <i>target</i>	110
<i>Input</i> di appoggio	110
<i>Input</i> de apoio	111
Istruzioni	111
Instruções	111
<i>Item</i>	111
Item	111
<i>Item</i> àncora	111
Item âncora	111
<i>Item</i> placebo	111
Item placebo	111
Opzioni	112
Opções	112

<i>Prompt</i>	112
<i>Prompt</i>	112
Punto di taglio	112
Escore de diferenciação	112
Scala di valutazione	113
Escala de avaliação	113

C. TIPOLOGIE

C. TIPOLOGIAS

TIPI DI VALUTAZIONE	115
TIPOS DE AVALIAÇÃO	115
Autovalutazione	115
Autoavaliação	115
Valutazione CLIL	116
Avaliação CLIL	116
Valutazione collettiva	116
Avaliação coletiva	116
Valutazione condivisa	117
Avaliação partilhada	117
Valutazione dinamica	117
Avaliação dinâmica	118
Valutazione a distanza	118
Avaliação à distância	118
Valutazione diffusa	118
Avaliação contínua	119

Valutazione formativa	119
<i>Avaliação formativa</i>	120
Valutazione informatizzata	120
<i>Avaliação informatizada</i>	120
Valutazione dell'interlingua	120
<i>Avaliação da interlíngua</i>	121
Valutazione tra pari	121
<i>Avaliação entre pares</i>	121
Valutazione personalizzata	122
<i>Avaliação personalizada</i>	122
Valutazione riferita alla norma	123
<i>Avaliação normativa</i>	123
Valutazione riferita a dei criteri	123
<i>Avaliação criterial</i>	123
Valutazione sommativa	124
<i>Avaliação somativa</i>	124
TIPDI ESAMI	124
<i>TIPOS DE EXAMES</i>	124
Portfolio	124
<i>Portfólio</i>	124
Prova per abilità	124
<i>Prova de habilidades</i>	124
Prova di accesso	125
<i>Prova de ingresso</i>	125

Prova adattiva	125
Prova adaptativa	125
Prova centralizzata	125
Prova externa	125
Prova di competenza	126
Prova de proficiência	126
Prova diagnostica	126
Prova diagnóstica	126
Prova di fine ciclo	126
Prova de conclusão	126
Prova integrata	126
Prova integrada	127
Prova di piazzamento	127
Prova de nivelamento	127
Prova di profitto	127
Prova de conhecimentos	128
TIPI DI ATTIVITÀ	128
TIPOS DE ATIVIDADES	128
Attività aperta	128
Atividade com resposta livre	128
Attività chiusa	129
Atividade com gabarito	129
Attività integrata	129
Atividade integrada	129

Attività di coppia	129
Atividade em dupla	129
Attività elegibile	129
Atividade com possibilidade de escolha	129
Attività di gruppo	129
Atividade em grupo	129
Attività individuale	129
Atividade individual	129
Attività di transcodificazione	130
Atividade de transcodificação	130
Attività semistrutturata	130
Atividade semiestruturada	130

D. PROPRIETÀ

D. PROPRIEDADES

PROPRIETÀ DI UNA PROVA/ ATTIVITÀ	131
PROPRIEDADES DE UMA PROVA/ATIVIDADE	131
Accettazione	131
Aceitação	131
Adeguatezza	131
Pertinência	131
Affidabilità	132
Confiabilidade	133
Autenticità	135
Autenticidade	135

Calibrazione	136
Adequação	136
Completezza	136
Exaustividade	136
Convergenza	136
Convergência	136
Copertura	136
Abrangência	137
Discriminatività	137
Discriminação	137
Durata	137
Duração	137
Economia	137
Economia	138
Elicitazione	138
Comportamento específico gerado por uma atividade	138
Estensione	139
Tamanho	139
Familiarità	140
Familiaridade	140
Impatto	140
Impacto	140
Praticabilità	140
Praticidade	140

Sicurezza	140
Segurança	140
Specificità	141
Especificidade	141
Trasparenza	141
Transparência	142
Validità	142
Validade	143
PROPRIETÀ DI <i>PROMPT</i> , ISTRUZIONI, <i>ITEM</i>	144
PROPRIEDADES DE <i>PROMPTS</i> , INSTRUÇÕES, <i>ITENS</i>	14
Accessibilità	144
Compreensibilidade	144
Calibrazione	144
Adequação	144
Discriminatività	144
Discriminação	145
Concisione	145
Concisão	145
Generatività	145
Produtividade	145
Indipendenza	145
Independência	145
Omogeneità	146
Homogeneidade	146

Specificità	146
Especificidades	146

E. OPERAZIONI

E. PROCEDIMENTOS OPERACIONAIS

Allestimento degli spazi	147
Organização do espaço	147
Allineamento dei giudizi	147
Convergência nas avaliações	148
Assegnazione del punteggio	148
Atribuição da pontuação	148
Comunicazione dei risultati	148
Divulgação dos resultados	148
Formattazione	149
Diagramação da prova	149
Generalizzazione	149
Generalização	149
Manipolazione del testo	149
Reelaboração do texto	149
Mappatura del testo	149
Mapeamento do texto	149
Ponderazione	150
Peso	150
<i>Pre-testing</i>	150
Pré-testagem	150

<i>Post-testing</i>	151
Pós-testagem	151
Recupero	151
Prova de recuperação	151
Revisione	151
Controle e revisão das questões de prova	151
Scelta del testo	151
Escolha do texto	15
Scelta del <i>format</i>	152
Escolha do tipo de prova	152
Scrittura degli <i>item</i>	152
Redação dos itens	152
Scrittura del testo	152
Redação do texto	153
Simulazione	153
Simulado	153
Somministrazione	153
Aplicação	153
<i>Standard Setting</i>	154
<i>Standard Setting</i>	154
Sviluppo di un testo orale a partire da una traccia	154
Preparação com base em esquema do texto oral de prova	154

F. ISTITUZIONI E INDIVIDUI

F. PESSOAS E INSTITUIÇÕES

Centro d'esame	155
Ente credenciado	155
Centro Certificatore	155
Ente certificador	155
Correttore	155
Corretor de redação	156
Esaminatore	156
Examinador	156
Falso positivo	157
Falso positivo	157
Falso negativo	157
Falso negativo	157
Giudice	157
Juiz	157
<i>Inlier</i>	157
<i>Inlier</i>	157
Intervistatore	157
Entrevistador	157
<i>Item Writer</i>	158
Redator de itens	158
Moderatore	158
Moderador	158

<i>Outlier</i>	158
<i>Outlier</i>	158
<i>Test taker</i>	158
Examinando	158

G. CRITICITÀ

G. ASPECTOS CRÍTICOS

Cattivi usi	159
Uso inappropriado	160
Difetti di progettazione	161
Erros de elaboração	161
Difetti del giudizio	161
Erros de avaliação	162
Difetti della scala	163
Escala inadequada	163
Difetti nella definizione del costrutto operativo	163
Erro na definição do constructo operacional	164

H. MISURAZIONI, RAPPRESENTAZIONI, VALORI

H. MEDIÇÕES, REPRESENTAÇÕES E VALORES

Analisi dei distrattori	167
Análise dos distratores	167
Analisi fattoriale	167
Análise fatorial	167

Analisi degli <i>item</i>	168
<i>Análise dos itens</i>	168
Analisi Rasch	168
<i>Análise Rasch</i>	169
Affidabilità (indice)	170
<i>Confiabilidade (índice)</i>	172
Asimmetria (coefficiente)	173
<i>Assimetria (coeficiente)</i>	174
Attrattività (indice)	175
<i>Capacidade de atração do distrator (índice)</i>	175
Comportamento differenziato degli <i>item</i> (DIF)	175
<i>Comportamento diferenciado dos itens (DIF)</i>	176
Correlazione (indice)	177
<i>Correlação (índice)</i>	178
Covarianza (indice)	179
<i>Covariância (índice)</i>	179
Curtosi (coefficiente)	179
<i>Curtose (coeficiente)</i>	180
Discriminatività (indice e correlazione Punto biseriale)	180
<i>Discriminação de elementos (índice e coeficiente de correlação ponto-biserial)</i>	182
Devianza	184
<i>Desvio</i>	184
Deviazione standard	184
<i>Desvio padrão</i>	184

Distribuzione	184
Distribuição	185
Errore di misurazione	185
Erro de medida	185
Facilità (indice)	186
Facilidade (índice)	187
Grafico a barre	188
Gráfico de barra	188
Grafico a dispersione	188
Gráfico de dispersão	189
Indagini qualitative	189
Pesquisas qualitativas	190
Indici di conformità (<i>Fit Statistics</i>)	190
Índices de conformidade (<i>Fit Statistics</i>)	192
Intervallo	195
Intervalo	195
Istogramma	195
Histograma	195
<i>Logit</i>	195
<i>Logit</i>	195
Media	195
Média	196
Mediana	196
Mediana	196

Moda	196
Moda	196
Normalità	196
Normalidade	197
Punteggio zeta (ζ score)	198
Escore-z (ζ score)	199
Significatività statistica	199
Significância estatística	201
Stima dell'effetto (<i>effect size</i>)	203
Tamanho do efeito (<i>effect size</i>)	203
Variabile	204
Variável	205
Varianza	206
Variância	206

I. METAVALUTAZIONE

I. META-AVALIAÇÃO

Affaticamento (<i>cognitive fatigue</i>)	207
Fadiga (<i>cognitive fatigue</i>)	207
Ansia da esame (<i>test anxiety</i>)	207
Estresse pré-prova (<i>test anxiety</i>)	208
Atteggiamento	208
Atitude perante a prova	208
Attribuzione causale	209
Atribuição de causalidade	209

Autoregolazione	210
<i>Autorregulação</i>	210
Controvalutazione	210
<i>Antiavaliação</i>	210
Convinzione	211
<i>Convicção</i>	211
<i>Effectance</i>	212
<i>Effectance</i>	212
Etica (della valutazione)	213
<i>Ética (da avaliação)</i>	213
Ipervalutazione	213
<i>Excesso de avaliações</i>	214
Motivazione	214
<i>Motivação</i>	215
Stile di apprendimento	215
<i>Estilo de aprendizagem</i>	215
Valori	216
<i>Valores</i>	216
Valutazione dei valutatori	216
<i>Avaliação dos avaliadores</i>	216

PARTE III. DOCUMENTI E RISORSE
PARTE III. DOCUMENTOS E RECURSOS

DOCUMENTI EUROPEI
DOCUMENTOS EUROPEUS

<i>Companion</i>	219
<i>Companion</i>	219
Portfolio Europeo per le Lingue	219
Portfólio Europeu de Línguas	220
Profili di varie lingue	220
Referenciais de várias línguas	220
Quadro Comune Europeo di Riferimento per le Lingue	221
<i>Quadro Europeu Comum de Referência para as Línguas</i>	221

RISORSE
RECURSOS

Risorse ALTE	223
Recursos ALTE	223
Risorse British Council	223
Recursos British Council	223
Risorse EALTA	224
Recursos EALTA	224
Risorse IATEFL TEASIG	224
Recursos IATEFL TEASIG	224

Risorse ILTA	224
Recursos ILTA	225
Risorse INTO EUROPE Project	226
Recursos INTO EUROPE Project	227
Risorse LAALTA	227
Recursos LAALTA	227
Risorse LTRG Lancaster	227
Recursos LTRG Lancaster	227
Riviste specializzate	228
Revistas especializadas	228
Sito del prof. Gianfranco Porcelli	229
Site do prof. Gianfranco Porcelli	229

PARTE IV. APPENDICI E BIBLIOGRAFIA

PARTE IV. APÊNDICES E BIBLIOGRAFIA

Appendice 1. Griglie analitiche per la valutazione della produzione linguistica	233
Apêndice 1. Tabelas analíticas para a avaliação da produção linguística	238
Appendice 2. Griglia di analisi di un testo audio (livello B2)	243
Apêndice 2. Tabela de análise de um texto de escuta (nível B2)	245
Appendice 3. Griglia semplificata di valutazione di un testo argomentativo	247
Apêndice 3. Tabela de avaliação simplificada de um texto argumentativo	249

Appendice 4. Generatività. Approfondimento	251
Apêndice 4. Produtividade. Aprofundamento	252
Appendice 5. Sottorappresentazione del costrutto. Approfondimento	253
Apêndice 5. Sub-representação do constructo. Aprofundamento	254
Appendice 6. Variabile esterna. Approfondimento	255
Apêndice 6. Variável externa. Aprofundamento	256
Riferimenti bibliografici	257
Referências bibliográficas	257

Prefazione dell'autore

Questo libro costituisce una mappa utile a quanti si avvicinano alla valutazione linguistica: insegnanti, ricercatori, studenti e amministratori. È diviso in due parti. La prima, dal tono discorsivo, definisce le coordinate fondamentali della valutazione linguistica secondo una prospettiva comunicativa, centrata sull'esercizio delle abilità linguistiche. La seconda, più analitica, è una compilazione di termini salienti della valutazione linguistica; il lettore vi potrà incontrare concetti noti, e altri meno.

Nello stilare l'elenco ci siamo ispirati a principi di parsimonia e chiarezza: abbiamo elaborato definizioni essenziali, che risultassero trasparenti anche al non addetto ai lavori. Una particolare cura è stata riservata alle questioni legate alla statistica, fornendo, laddove opportuno, dei valori di riferimento per i singoli indici, in funzione dei diversi contesti valutativi. In sintonia con testi analoghi prodotti in terra inglese (quali DAVIES *et al.* 1999; COOMBE 2018 [di libero accesso]; ALTE 1999 [testo multilingue]), il glossario è stato redatto nell'ottica di promuovere buone pratiche, nella consapevolezza delle numerose variabili che incidono sulla qualità di un test.

Paolo Torresan
Università Federal Fluminense
Rio de Janeiro, maggio 2021

Prefácio do autor

Este livro é um mapa útil para quem se aproxima da avaliação linguística: professores, pesquisadores, estudantes e gestores. Divide-se em duas partes. A primeira, com um tom discursivo, define as coordenadas fundamentais da avaliação linguística, seguindo uma perspectiva comunicativa que está centrada no exercício das habilidades linguísticas. A segunda, mais analítica, constitui uma coletânea dos termos mais importantes da avaliação linguística; o leitor vai encontrar conceitos familiares, e outros menos.

Os princípios de parsimônia e de clareza nos guiaram na preparação da lista: elaboramos definições que resultassem transparentes para o público em geral. Foi dada atenção especial às questões relacionadas com a estatística, fornecendo, sempre que necessário, valores de referência para cada índice, em função dos diferentes contextos de avaliação. Em consonância com textos semelhantes produzidos em inglês (como DAVIES *et al.* 1999; COOMBE 2018 [de livre acesso]; ALTE 1999 [texto multilíngue]), o glossário foi elaborado com o objetivo de promover boas práticas, estando consciente das muitas variáveis que afetam a qualidade de um teste.

Paolo Torresan
Universidade Federal Fluminense
Rio de Janeiro, maio de 2021

Introduzione

Il glossario bilingue messo a punto da Paolo Torresan non è una semplice raccolta di vocaboli in versione italiano/portoghese sulla tematica della valutazione linguistica. Questo volume è una vera propria mappa per orientarsi in un ambito interdisciplinare che spazia dalla valutazione linguistica alla metodologia della valutazione, dalla misurazione e metrologia alle dimensioni statistiche e di utilizzazione di pacchetti informatici dedicati. Le circa 200 voci che lo compongono offrono un reticolo concettuale che, accanto a una consultazione rapida e a rimandi circolari, garantisce un'esaustiva trattazione e un arricchente piano per l'autoformazione professionale.

Chiunque volesse avvicinarsi alla valutazione linguistica in qualità di insegnante, ricercatore, studente o amministratore di istituzioni scolastiche può ritenere il glossario un valido strumento per la comprensione di tale ambito di ricerca e di studio.

In particolare, il glossario torna utile alle figure professionali impegnate nella progettazione di strumenti per la valutazione linguistica: il costruttore di formati valutativi (*l'item writer*), da un lato, e i responsabili di centri o strutture per la predisposizione e analisi dei risultati di prove di accertamento, di certificazione e di valutazione, dall'altro.

Dall'indice si può riconoscere l'impianto operativo del glossario.

Nella **prima parte** si delineano le principali coordinate della valutazione linguistica, attraverso un'articolata e aggiornata prospettiva comunicativa. Va riconosciuto come la valutazione linguistica si confronta, infatti, non solo con la questione dei formati, vale dire delle tecniche, della dimensione operativa, ma anche con quella dei fini e delle funzioni della valutazione. Il che cosa sia "oggetto" della valutazione e il "perché" della valutazione sono i pilastri e, al medesimo tempo, gli

orizzonti che guidano le scelte e le modalità stesse della valutazione, in ogni campo. Conoscenze o abilità? Comprensione di quali tipi di testo? Produzione di quali tipi di testo? Sono queste alcune delle piste di riflessione che l'autore pone in testa al glossario, perché è proprio dalle riflessioni teoriche, dai quadri di riferimento della lingua, dalla visione stessa delle funzioni linguistiche e comunicative che devono discendere le scelte operative e strumentali.

Le diverse Sezioni del glossario (**seconda parte**) sono ordinate per tematiche. La nostra sintetica introduzione ripercorre la catalogazione proposta dall'autore.

I chiarimenti sull'interdipendenza tra teoria e prassi, tra mezzi e fini, tra funzioni ed effetti della valutazione portano il glossario a concentrarsi sui termini salienti della valutazione linguistica (Sezione A). La valutazione linguistica può riferirsi a diversi ambiti:

- a) le *conoscenze linguistiche* (es. lessico, grammatica, fonologia, ortografia);
- b) le *abilità linguistiche* (leggere, ascoltare, parlare, scrivere);
- c) le *abilità di mediazione* (inerenti a operazioni interlinguistiche o intralinguistiche).

Le diverse competenze oggetto di valutazione sono proprio le coordinate che definiscono i “domini” e i “costrutti” e compongono i “sillabi”.

E come distinguere, analizzare e costruire le parti costitutive di una prova linguistica? (Sezione B). Attraverso le definizioni di “descrittori” e dei formati di “item” si introducono le analisi dei “testi” e delle “performance”. Sono proprio le prestazioni che gli allievi offrono in risposta agli stimoli proposti che permettono di costruire scale e griglie per la valutazione, e successivamente di definire i livelli di sufficienza (i cosiddetti punti di taglio).

Non è, comunque, solo l'oggetto a determinare le differenze in seno alla valutazione linguistica. La valutazione varia anche in funzione di altri parametri (Sezioni Ca), quali: i tempi e le funzioni (valutazione formativa, dinamica, continua, normativa, criteriale, sommativa), le progettazioni didattiche (valutazione CLIL, dell'interlingua) e le modalità (valutazione individuale, collettiva, condivisa, a distanza, informatizzata, tra pari, personalizzata). Le diverse accezioni individuano azioni e strumenti distinti, proprio perché rimandano a specificità didattiche, livelli di prestazioni e piani valutativi differenti.

L'autore passa in rassegna, successivamente, i tipi diversi di esami e di prove (Sezione Cb), strutturati secondo tecniche specifiche (Sezione Cc).

Nella Sezione D il glossario affronta le proprietà che ogni formato valutativo dovrebbe possedere e che qualsiasi responsabile della pianificazione di prove per l'accertamento e la certificazione delle abilità linguistiche dovrebbe garantire. Da un lato, le caratteristiche peculiari di prove/attività di valutazione (Da) e, dall'altro, quelle degli *item*, delle istruzioni e dei *prompt* (Db). Si fa qui riferimento a concetti la cui definizione garantisce quel rigore metodologico necessario ai diversi piani della valutazione.

Con la Sezione E si dettagliano le procedure relative alla somministrazione delle prove. Le diverse operazioni – dall'allestimento degli spazi all'assegnazione dei punteggi e loro comunicazione – sono indispensabili per la credibilità, la funzionalità e il valore didattico che le dimensioni valutative possono assumere nei contesti sia scolastici, che di certificazione linguistica. Strettamente connessa, la Sezione successiva (F) illustra quali sono le figure professionali preposte alla valutazione, soprattutto sul versante della certificazione. I Centri d'esame e di certificazione sono oggi chiamati a svolgere un'azione sia di formazione che di attestazione rigorosa dei livelli linguistici, nel rispetto dei quadri di riferimento e delle politiche educative.

Giudici e esaminatori sono difatti i responsabili e garanti delle proprietà che ogni valutazione deve possedere, illustrate nella Sezione D.

L'autore presenta, inoltre, le criticità che possono dipendere da difetti di progettazione o di dimensionamento del costruito operativo (Sezione G). Ricca e decisamente più tecnica è poi la Sezione dedicata alla statistica (H). Le voci offrono riferimenti alla pluralità di analisi, indici statistici e livelli di significatività connessi, largamente utilizzati dagli specialisti, nonché alle modalità di rappresentazione grafica dei risultati. Segue una Sezione legata ad aspetti di metavalutazione (Sezione I). In essa si affrontano questioni, quali l'ansia da esame, la motivazione, l'affaticamento, gli stili di apprendimento dei candidati, ecc., che possono interferire sui livelli di prestazione e che quindi devono essere considerate con estrema attenzione.

Il glossario si chiude con due parti ulteriori: nella **terza** si indicano risorse e materiali utili all'approfondimento; la **quarta parte**, infine, comprende una serie di appendici e una bibliografia minima.

Superfluo sottolineare, dopo questa carrellata, il valore conoscitivo e di approfondimento che il glossario offre non solo agli addetti del settore. Di fatto, è difficile organizzare la complessità lessicale e concettuale di un ambito professionale e di processi complessi, multisetoriali e interdisciplinari. Paolo Torresan è riuscito largamente in questo intento, fornendo un documento plurilingue, che apre a confronti di ordine disciplinare. Lavorare sul *testing* e sulle diverse dimensioni della valutazione linguistica non può e non deve essere appannaggio solo di tecnici e specialisti del settore, ma deve interessare tutte le istituzioni e tutti i soggetti coinvolti, nel rispetto della pluralità delle dimensioni sociopolitiche e culturali e in una prospettiva globale e internazionale.

Guido Benvenuto

Università degli Studi di Roma "La Sapienza"

Introdução

O glossário bilingüe elaborado por Paolo Torresan não é uma simples coletânea de palavras nas versões italiano/português sobre a avaliação linguística. Este livro é um verdadeiro mapa para se orientar em um contexto interdisciplinar que abrange desde a avaliação linguística até a metodologia de avaliação, desde a medição e metrologia até as dimensões estatísticas e de uso de pacotes especializados de informática. Os cerca de 200 verbetes que o compõem oferecem uma rede conceitual que, além de uma consulta rápida e de referências circulares, garante um tratamento exaustivo e um projeto enriquecedor para a autoformação profissional.

Qualquer um que queira se aproximar da avaliação linguística, seja professor, pesquisador, estudante ou administrador de instituições de ensino, pode considerar o glossário uma ferramenta valiosa para a compreensão deste campo de pesquisa e de estudo.

O glossário é útil, principalmente, para os profissionais envolvidos na concepção de ferramentas de avaliação linguística: por um lado, o construtor de formatos avaliativos (o redator de itens); por outro, os chefes de centros ou estruturas para a apuração e análise dos resultados dos diferentes tipos de avaliação.

A partir do índice, podemos reconhecer o sistema operacional do glossário.

A **primeira parte** delinea as principais coordenadas da avaliação linguística, através de uma perspectiva comunicativa articulada e atualizada. Há que reconhecer que a avaliação linguística se defronta não só com a questão dos formatos, ou seja, das técnicas, da dimensão operacional, mas também com a questão dos objetivos e das funções da avaliação. O que é "objeto" da avaliação e o "porquê" da avaliação são os pilares e, ao mesmo tempo, os horizontes que norteiam as escolhas e as

modalidades da avaliação, em todos os campos. Conhecimentos ou habilidades? Compreender que tipos de texto? Produzir que tipos de texto? Estas são algumas das linhas de reflexão que o autor destaca no início do glossário, porque é precisamente a partir das reflexões teóricas, a partir dos quadros de referência da língua, da própria visão das funções linguísticas e comunicativas que devem derivar as escolhas operativas e instrumentais.

As diferentes seções do glossário (**segunda parte**) estão organizadas por temáticas. Nossa breve introdução percorre a catalogação proposta pelo autor.

Esclarecimentos sobre a interdependência entre teoria e prática, entre meios e fins, entre funções e efeitos da avaliação levam o glossário a se concentrar nos termos salientes da avaliação linguística (seção A). A avaliação linguística pode abranger vários domínios:

- a) os conhecimentos linguísticos (ex. léxico, gramática, fonologia, ortografia);
- b) as habilidades linguísticas (ler, ouvir, falar, escrever);
- c) as habilidades de mediação (relacionadas com operações interlinguísticas ou intralinguísticas).

As diferentes competências a serem avaliadas são exatamente as coordenadas que definem os "domínios" e os "constructos" e compõem os "componentes curriculares". E como distinguir, analisar e construir as partes que constituem qualquer prova de avaliação linguística? (seção B). Através das definições de "descritores" e dos formatos de "itens", é introduzida a análise dos "textos" e das "performances". É precisamente o desempenho que os alunos oferecem em resposta aos estímulos propostos que permite construir escalas e tabelas para a avaliação, e depois definir os níveis de suficiência (os chamados escores de diferenciação).

No entanto, não é apenas o objeto que determina as diferenças na avaliação linguística. A avaliação também varia de acordo com outros parâmetros (seção Ca),

tais como: tempos e funções (avaliação formativa, dinâmica, contínua, normativa, criterial, somativa), o tipo de escolha metodológica (avaliação CLIL, avaliação da interlíngua) e modalidades (avaliação individual, coletiva, partilhada, à distância, informatizada, entre pares, personalizada). As diferentes aceções identificam diferentes ações e ferramentas, justamente por se referirem a especificidades educativas, diferentes níveis de desempenho e programas de avaliação.

O autor então revisa os diferentes tipos de exames e provas (seção Cb), estruturados de acordo com técnicas específicas (seção Cc).

Na seção D, o glossário aborda as propriedades que cada formato de avaliação deveria possuir e que qualquer pessoa responsável pelo planejamento das provas para a verificação e proficiência das habilidades linguísticas deveria garantir. Por um lado, as características específicas das provas/atividades de avaliação (Da) e, por outro, as características dos itens, das instruções e dos *prompts* (Db). Esta seção se refere a conceitos cuja definição garante o rigor metodológico necessário para os diferentes programas de avaliação.

A seção E especifica os procedimentos de aplicação das provas. As diferentes operações - desde a organização do espaço até a atribuição da pontuação e a divulgação dos resultados - são indispensáveis para a credibilidade, a funcionalidade e o valor didático que as dimensões avaliativas podem assumir tanto no contexto escolar, quanto no de proficiência linguística. Intimamente interligada, a próxima seção (F) ilustra quais figuras profissionais são responsáveis pela avaliação, especialmente em termos de proficiência. Os entes credenciados e certificadores são agora chamados a realizar a formação e a certificação rigorosa dos níveis linguísticos, em conformidade com os quadros de referência da língua e com as políticas educacionais. Juízes e examinadores são, de fato, os responsáveis e garantes das propriedades que cada avaliação deve possuir, e que estão ilustradas na Seção D.

O autor também apresenta os aspectos críticos que podem resultar de erros de elaboração ou de adequação do constructo operacional (Seção G). Rica e decididamente mais técnica é então a seção dedicada à estatística (H). Os verbetes remetem à pluralidade de análises, índices estatísticos e níveis de significância relacionados, que são amplamente utilizados por especialistas, inclusive a representação gráfica dos resultados. Em seguida, há uma seção vinculada aos aspectos de meta-avaliação (Seção I). Ela aborda questões como estresse pré-prova, motivação, fadiga, estilos de aprendizagem dos candidatos, dentre outros, que podem interferir nos níveis de rendimento e que, portanto, devem ser considerados com muito cuidado.

O glossário encerra com mais duas partes: na **terceira parte**, recursos e materiais são indicados para estudo posterior; a **quarta parte**, por fim, inclui uma série de apêndices e uma bibliografia mínima.

Será supérfluo frisar, depois deste apanhado, o valor do glossário em termos de conhecimento e de aprofundamento para os profissionais do setor e outros interessados. De fato, é difícil organizar a complexidade lexical e conceitual de um campo profissional e de processos complexos, multissetoriais e interdisciplinares. Paolo Torresan conseguiu com sucesso este objetivo, fornecendo um documento multilíngue, que atinge diferentes áreas de conhecimento. Trabalhar com a avaliação linguística e suas diferentes dimensões não pode e não deve ser prerrogativa apenas de técnicos e especialistas na área, mas deve envolver todas as instituições e todas as partes interessadas, respeitando a pluralidade de dimensões sociopolíticas e culturais e em uma perspectiva global e internacional.

Guido Benvenuto

Università degli Studi di Roma “La Sapienza”

Parte I
Coordinate fondamentali

Parte I
Coordenadas fundamentais

Cosa significa valutare?

Valutare significa esprimere un *giudizio di valore*. Nella vita di tutti i giorni ci capita spesso di valutare: l'efficienza di un servizio, la qualità di un prodotto, i comportamenti delle persone che ci circondano e i nostri stessi comportamenti.

Ogni forma di valutazione presuppone un'osservazione, un oggetto di osservazione e un parametro di giudizio. Il giudizio può essere

- *semplice*; si concretizza in distinzioni polari (bello/brutto; positivo/negativo; giusto/sbagliato, ecc.) o tutt'al più opera su scale limitate (scarso/insufficiente/sufficiente/buono/ottimo);
- *complesso*, con la triangolazione di più dati (valutare se cambiare lavoro, casa o città, per esempio, richiede di soppesare una serie di pro e di contro).

O que significa avaliar?

Avaliar significa fazer um *juízo de valor*. Na vida cotidiana avaliamos muitas vezes: a eficiência de um serviço, a qualidade de um produto, o comportamento das pessoas ao nosso redor e o nosso próprio comportamento.

Toda forma de avaliação requer uma observação, um objeto de observação e um parâmetro de julgamento. O julgamento pode ser

- *simples*, é expresso através de oposições (bonito/feio; positivo/negativo; certo/errado, etc.) ou, quando muito, trabalha com escalas limitadas (ruim/insuficiente/suficiente/bom/excelente);
- *complexo*, com a triangulação de mais dados (avaliar se mudar de emprego, casa ou cidade, por exemplo, exige que você pondere uma série de prós e contras).

Contesti valutativi

Esistono diversi tipi di valutazione. La *valutazione* può essere *formale*, se realizzata in un preciso momento, mediante test programmati, o *informale*, se prescinde da prove allestite *ad hoc*, per basarsi piuttosto su impressioni e appunti aneddotici raccolti nel tempo.

Si tende a distinguere, inoltre, una *valutazione sommativa* da una *valutazione formativa*. Mentre la prima si concretizza nel voto/giudizio formulato al termine di un percorso di studio, la seconda ha lo scopo di formare nello studente una consapevolezza circa la qualità della propria prestazione. In altre parole, la *valutazione formativa* conta su un *feedback* continuo, mediante il quale l'allievo prende coscienza dei punti di forza e degli aspetti da migliorare, e, in caso, delle strategie da intraprendere.

La *valutazione formativa* è importante in ambito scolastico, considerato che sposta l'accento dalla prestazione all'alunno. L'obiettivo, dal punto di vista di chi apprende, non è, in quest'ottica, tanto prendere un buon voto, quanto sentirsi protagonista dell'evoluzione della propria competenza: cosa può fare per migliorare? (cfr. PANADERO 2018; TAI *et. al.*, 2018). In ambito extrascolastico, la *valutazione formativa* contraddistingue il modo in cui si consolidano le abilità manuali: il maestro artigiano non somministra dei test agli apprendisti, ma restituisce loro delle indicazioni affinché ciascuno lavori al meglio delle proprie possibilità. Gli standard, in questo caso, si interiorizzano, si affina l'attenzione rispetto alle strategie implementabili e la motivazione si rafforza ad ogni minimo traguardo raggiunto (cfr. BORKOWSKI, MATHUKRISHNA, 1992).

Contextos de avaliação

Existem diferentes tipos de avaliação. A *avaliação* pode ser *formal*, realizada em um momento específico, por meio de testes programados, ou *informal*, quando dispensa provas criadas *ad hoc*, e se fundamenta mais em impressões e anotações significativas coletadas ao longo do tempo.

Tende-se a distinguir também uma *avaliação somativa* de uma *avaliação formativa*. Enquanto a primeira se efetiva através de nota/avaliação formulada no final de um percurso de estudo, a segunda pretende treinar no estudante uma consciência sobre a qualidade do seu próprio desempenho. Em outras palavras, a *avaliação formativa* consiste em um *feedback* contínuo, através do qual o aluno toma consciência dos pontos fortes e dos aspectos a melhorar, e, se necessário, das estratégias a adotar.

A *avaliação formativa* é importante nas escolas, pois desloca o foco da avaliação para o aluno. O objetivo, do ponto de vista de quem aprende nesta perspectiva, não é só tirar uma nota boa, mas também se sentir protagonista da evolução da própria competência: o que o aluno pode fazer para melhorar? (cfr. PANADERO 2018; TAI *et. al.*, 2018). No âmbito extraescolar, a *avaliação formativa* distingue o modo como as habilidades manuais são consolidadas: o mestre artesão não aplica testes aos aprendizes, mas dá instruções para que cada um trabalhe da melhor forma possível. Os padrões, neste caso, são internalizados, a atenção com relação às estratégias implementadas são aprimoradas e a motivação é reforçada com cada meta mínima alcançada (cfr. BORKOWSKI, MATHUKRISHNA, 1992).

Tipi di prova

Intendiamo per *prove* (o esami) un pacchetto di attività, ciascuna delle quali è volta ad accertare un aspetto particolare della conoscenza/competenza. La prova (o l'esame) ENEM o la prova (o l'esame) CELPE-Bras, per esempio, sono costituite da una serie articolata di attività.

Le prove più diffuse in ambito educativo sono le prove di profitto (*achievement test*). Si tratta di prove somministrate secondo una certa scadenza, l'obiettivo delle quali consiste nel verificare quanto lo studente ha appreso in un certo arco di tempo; hanno un carattere prevalentemente *sommativo* (queste stesse prove, peraltro, possono informare, indirettamente, sulla qualità dell'insegnamento: se in una classe più del 50% degli allievi non ha assimilato una serie di contenuti, vuol dire che l'insegnante non ha insegnato bene).

Altro genere sono le *prove di piazzamento* (*placement test*), le quali vengono somministrate prima di un percorso di studio e hanno una funzione di livellamento: distribuire gli allievi in classi che corrispondono a diversi livelli di competenza.

Le *prove di accesso*, altro caso ancora, hanno invece una funzione selettiva in riferimento a un certo percorso di studio a cui uno voglia accedere (come avviene nella prova ENEM) o a una professione che uno voglia esercitare.

Infine, le *prove di competenza* (*proficiency test*), elaborate da un centro certificatore, si rivolgono a un vasto bacino di utenza e hanno per scopo l'accertamento della competenza, a prescindere dal percorso di studio del singolo candidato (cfr. VEDOVELLI, 2005).

Tipos de prova

Entende-se por *provas* (ou exames) um conjunto de atividades, cada uma delas é destinada a verificar um aspecto particular do conhecimento/da competência. O exame ENEM e o exame CELPE-Bras, por exemplo, consistem em um conjunto articulado de atividades.

As provas mais comuns no âmbito escolar são as provas de conhecimento (*achievement test*). São provas aplicadas de acordo com um determinado prazo, cujo objetivo é verificar o que o estudante aprendeu durante um determinado período de tempo; as provas têm um caráter predominantemente *somatório* (no entanto, estas mesmas provas podem informar indiretamente sobre a qualidade do ensino: se em uma turma mais de 50% dos alunos não assimilou uma série de conteúdos, isso significa que o professor não ensinou bem).

Outro tipo são as *provas de nivelamento* (*placement test*), que são aplicadas antes de um percurso de estudo e têm função de nivelamento: distribuir os alunos em turmas que correspondem a diferentes níveis de competência.

As *provas de ingresso*, ainda outro caso, têm uma função seletiva em relação a um determinado percurso de estudo que se quer acessar (como é o caso do exame ENEM) ou a uma profissão que se quer exercer.

Por último, as *provas de proficiência* (*proficiency test*), que são elaboradas por um ente certificador, destinam-se a um grande público e têm por objetivo determinar a competência, seja qual for o percurso de estudo de cada candidato (cfr. VEDOVELLI, 2005).

Effetti della valutazione

Qualsiasi giudizio esercita un impatto, sia in termini pratici che psicologici, sulla vita di un individuo. Valutare, in questo senso, è un'operazione delicatissima, durante la quale il valutatore rischia di etichettare l'apprendente sulla base dell'esito della prestazione. Apprendenti ansiosi non danno, in effetti, il meglio di sé in sede d'esame; a un risultato negativo può seguire uno scoramento a cui si lega una maggiore apprensione in occasione di un nuovo test. Si creano, così, spirali negative, le ripercussioni delle quali sulla motivazione possono portare alla disaffezione verso la materia o, addirittura, all'abbandono del corso. Del resto, pure giudizi eccessivamente generosi, che tendono a livellare verso l'alto un'intera classe, non rendono giustizia alla stima delle competenze: se tutto è perfetto, brillante, magnifico, come distinguere ciò che buono da ciò che è meno buono? L'insegnante troppo 'generoso' rischia di apparire, oltretutto, poco preparato agli occhi di chi impara.

Efeitos da avaliação

Qualquer avaliação tem um impacto, tanto em termos práticos quanto em termos psicológicos, na vida de um indivíduo. Avaliar, nesse sentido, é uma operação muito delicada, durante a qual o avaliador corre o risco de rotular o aluno, a partir do resultado da avaliação. Na verdade, alunos ansiosos não dão o melhor de si na prova; um resultado negativo pode gerar um desencorajamento e uma maior apreensão no momento de um novo teste. Desta forma, são criadas espirais negativas, cujas repercussões na motivação podem levar ao desinteresse pela matéria ou, até mesmo, ao abandono do curso. Além disso, avaliações excessivamente generosas, que tendem a nivelar por cima uma turma inteira, não refletem uma avaliação eficaz: se tudo é perfeito, brilhante, magnífico, como distinguir o que é bom do que é menos bom? Aliás, o professor ‘generoso’ demais corre o risco de parecer mal preparado aos olhos de quem aprende.

Didattica e ricerca

La valutazione è un aspetto della competenza professionale trascurato nei corsi di formazione. Si presume che valutare sia relativamente facile, posto che l'insegnante domini la disciplina e somministri delle prove attenendosi a una certa scala. Pur si riconosce che esiste un ambito – quello docimologico – fatto di calcoli complessi, relativo agli esami applicati su larga scala, ma lo si ritiene di dominio di esperti.

La valutazione appare così come una disciplina che può assumere connotazioni opposte: intuitiva, da un lato, e molto complessa dall'altro.

La nostra convinzione è che esista una zona intermedia, tra la didattica e la ricerca, che può essere condivisa. Ed è anzi un bene che i due mondi – didattica e ricerca – dialoghino tra loro. Occorre, in effetti, da un lato, che l'insegnante approfondisca questioni di cui il ricercatore è l'esperto:

- abbia una chiarezza sul costrutto di competenza (cosa significa saper scrivere? Saper leggere? ecc.);
- conosca le proprietà fondamentali di un test;
- si attenga alle buone pratiche in materia di redazione degli *item*.

Dall'altro lato, il ricercatore o l'*item writer* di professione dovrebbero ispirarsi ai nuovi orientamenti nell'ambito dell'educazione linguistica per formulare proposte originali e in linea con quello che succede effettivamente in classe, traendo vantaggio della conoscenza di tecniche e di strategie in possesso dall'insegnante.

Pensiero critico (del ricercatore e dell'*item writer*) e pensiero divergente (dell'insegnante) possono arrecare, insomma, un beneficio reciproco. Se, al

contrario, non si saldano i due ambiti – insegnamento e valutazione – la stessa coscienza dell'apprendente soffre di una lacerazione: a lezione ci si diverte, mentre in sede d'esame si soffre.

Ensino e pesquisa

A avaliação é um aspecto negligenciado da competência profissional nos cursos de formação. Supõe-se que avaliar seja relativamente fácil, à medida que o professor domina a disciplina e aplica as provas atendo-se a uma certa escala. Do mesmo modo, considera-se que existe uma área – a do estudo docimológico – feita de cálculos complexos, relacionada aos exames aplicados em larga escala, que é de domínio de especialistas.

A avaliação aparece como uma disciplina que pode assumir conotações opostas: intuitiva por um lado e muito complexa por outro.

A nossa convicção é de que existe uma área intermediária entre o ensino e a pesquisa, que pode ser partilhada. E é realmente uma coisa boa que os dois mundos – ensino e pesquisa – dialoguem um com o outro. Na verdade, o professor, por um lado, precisa investigar questões das quais o pesquisador é o especialista:

- precisa ter uma compreensão clara do constructo de competência (o que significa saber escrever? Saber ler? etc.);
- precisa conhecer as propriedades fundamentais de um teste;
- precisa seguir boas práticas na redação dos itens.

Por outro lado, o pesquisador e o redator profissional de itens devem se inspirar nas novas diretrizes no campo da educação linguística para formular propostas originais e em consonância com o que realmente acontece em sala de aula, beneficiando-se do conhecimento do professor sobre técnicas e estratégias.

O pensamento crítico (do pesquisador e do redator de itens) e o pensamento divergente (do professor) podem trazer, em última análise, um

benefício mútuo. Se, ao contrário, as duas áreas – ensino e avaliação – não se juntam, a própria consciência do aluno sofre com uma ruptura: na aula você se diverte, enquanto no exame você sofre.

Valutazione linguistica: conoscenze vs abilità

Nell'ambito della valutazione linguistica, il giudizio può riferirsi a

- *conoscenze linguistiche* (es. lessico, grammatica, fonologia, ortografia);
- *abilità linguistiche* (leggere, ascoltare, parlare, scrivere);
- *abilità di mediazione* (riferibili a operazioni interlinguistiche o intralinguistiche).

Mentre il giudizio relativo alle *conoscenze linguistiche* è relativamente semplice, quello riferito alle *abilità linguistiche* è complesso, e ancor di più lo è quello riferito alle *abilità di mediazione*.

Per valutare quanto lo studente conosce la grammatica o il lessico della lingua oggetto di studio si allestiscono batterie di esercizi riferite a conoscenze dichiarative. L'oggetto di misurazione è circoscritto: le regole o il lessico che lo studente ha appreso all'interno di un certo percorso di studio. Si dirà allora che, sulla base degli esiti raggiunti in un test grammaticale o lessicale, l'apprendente ha una certa conoscenza del *passé composé* del francese, o del *simple past* dell'inglese, o del *passato prossimo* dell'italiano, o ancora del campo lessicale riferito alla casa, alle abitudini quotidiane, agli oggetti presenti in classe, ecc. (cfr. READ, 2000; PURPURA, 2004). Un discorso analogo può essere svolto in riferimento alle conoscenze fonoprosodiche (PENNINGTON, ROGERSON-REVELL, 2019).

Più complesso è stimare le *abilità linguistiche* e quelle di *mediazione*, dicevamo. In questi casi, il costrutto (cioè l'idea astratta della competenza oggetto di misurazione) è più ampio rispetto a quello delle *conoscenze linguistiche*. Non solo, è qualitativamente diverso. Le *conoscenze linguistiche* sono conoscenze

dichiarative, abbiamo detto. Le *abilità linguistiche* (e la *mediazione*, che sulle stesse poggia, e di cui parleremo più avanti), invece, sono forme di conoscenza tacita, procedurale. Il passaggio tra i due tipi di conoscenza non è automatico. Sapere la grammatica di una lingua o conoscere molte parole di una lingua non significa saper parlare o saper scrivere in quella lingua (RASTELLI, 2020). Le *abilità* si formano mediante l'uso comunicativo della lingua. Leggere un testo è qualcosa di più della decodifica degli enunciati: implica un'interpretazione coerente dei concetti veicolati dal testo. Scrivere, alla pari, non equivale a infilare una serie di frasi corrette, una dopo l'altra, ma prevede un'organizzazione delle idee, facendo riferimento a un destinatario e nell'ottica di raggiungere un certo scopo comunicativo.

Se in passato le *abilità* erano misurate in maniera indiretta, valutando il possesso di singole componenti (per esempio, si desumeva l'abilità di scrittura di un candidato sulla base del fatto che riuscisse a scrivere sotto dettatura, e quindi sul fatto che conoscesse, in sostanza, la corrispondenza fonema-grafema), oggi giorno all'interno di una prospettiva comunicativa (per la quale diventa centrale il fatto che lo studente sappia usare la lingua per raggiungere certi scopi comunicativi) le *abilità* sono misurate direttamente, considerando se e quanto l'apprendente è capace di agire strategicamente con la lingua (WEIR *et al.*, 2013). In una prospettiva di questo tipo può addirittura capitare che le conoscenze lessico-grammaticali dell'apprendente siano desunte a partire dalla sua *performance* linguistica, ovvero dalla sua abilità di produrre testi scritti o orali. Esistono, in tal senso, prove di certificazione linguistica, come il PLIDA per l'italiano, che si compongono di sole attività comunicative e non prevedono alcun esercizio grammaticale o lessicale.

Consideriamo a seguire una serie di indicazioni inerenti la valutazione delle *abilità comunicative* (*ricettive* e *produttive*), per affrontare in un momento successivo la questione della valutazione delle *abilità di mediazione*.

Avaliação linguística: conhecimentos vs. habilidades

No contexto da avaliação linguística, a avaliação pode referir-se a

- *conhecimentos linguísticos* (ex. léxico, gramática, fonologia, ortografia);
- *habilidades linguísticas* (ler, ouvir, falar, escrever);
- *habilidades de mediação* (relacionadas com operações interlinguísticas ou intralinguísticas).

Enquanto a avaliação das *competências linguísticas* é relativamente simples, a avaliação das *habilidades linguísticas* é complexa e, ainda mais, a das *habilidades de mediação*.

Para avaliar o quanto o estudante conhece a gramática ou o léxico da língua em estudo, prepara-se uma série de exercícios ligados a conhecimentos declarativos. A avaliação se refere a algo muito específico: trata-se de regras ou do léxico que o estudante aprendeu dentro de um determinado percurso de estudo. Em seguida, será dito que, com base nos resultados obtidos em um teste gramatical ou lexical, o aluno tem algum conhecimento do *passé composé* do francês, ou do *simple past* do inglês, ou do *passato prossimo* do italiano, ou do campo lexical ligado à casa, aos hábitos diários, aos objetos na sala de aula, etc. (cfr. READ, 2000; PURPURA, 2004). Um discurso semelhante pode ser feito em relação aos conhecimentos fonéticos e de prosódia (PENNINGTON, ROGERSON-REVELL, 2019).

Mais complexo é estimar *habilidades linguísticas* e *habilidades de mediação*, como dissemos anteriormente. Nesses casos, o constructo (isto é, a ideia abstrata da competência sendo medida) é mais amplo do que o constructo dos *conhecimentos linguísticos*. Não só isso, é qualitativamente diferente. Os *conhecimentos linguísticos* são

conhecimentos declarativos, conforme dissemos. As *habilidades linguísticas* (e a *mediação*, que nelas se baseia, e sobre a qual falaremos mais tarde), em vez disso, são formas de conhecimento tácito e processual. A transição entre os dois tipos de conhecimento não é automática. Conhecer a gramática de uma língua ou saber muitas palavras de uma língua não significa saber falar ou saber escrever nessa língua (RASTELLI, 2020). As *habilidades* são formadas pelo uso comunicativo da língua. Ler um texto é mais do que decodificar um enunciado: implica uma interpretação coerente dos conceitos veiculados pelo texto. Do mesmo modo, escrever não equivale a inserir uma série de frases corretas, uma após a outra, mas requer uma organização de ideias, fazendo referência a um destinatário e com o objetivo de alcançar um determinado propósito comunicativo.

Se no passado as *habilidades* eram medidas indiretamente, avaliando a aquisição de cada componente (por exemplo, a habilidade de escrita de um candidato era depreendida com base no que ele fosse capaz de escrever sob ditado, e, portanto, no fato de que ele conhecesse, basicamente, a correspondência entre grafemas e fonemas), hoje em dia, no âmbito da perspectiva comunicativa (na qual se torna central que aluno saiba como usar a língua para alcançar determinados objetivos comunicativos), as *habilidades* são medidas diretamente, considerando se e quanto o aluno é capaz de agir estrategicamente com o idioma (WEIR *et al.*, 2013). Em uma perspectiva deste tipo, pode até acontecer que os conhecimentos lexicais e gramaticais do aluno sejam inferidos a partir da sua *performance* linguística, ou seja, da sua habilidade em produzir textos escritos ou orais. Existem, neste sentido, exames de proficiência linguística, como o PLIDA para o italiano, que são compostos apenas por atividades comunicativas e não fornecem nenhum exercício gramatical ou lexical.

Consideramos, a seguir, uma série de indicações com relação à avaliação das *habilidades comunicativas* (*receptivas* e *produtivas*), para abordar posteriormente a questão da avaliação das *habilidades de mediação*.

Valutare la comprensione

Nello stimare le abilità di comprensione in possesso di un individuo, l'attenzione del valutatore è tenuta a operare su più piani:

- *il testo oggetto di comprensione;*
- *il tipo di attività;*
- *gli elementi (item) di cui una stessa attività si compone.*

Consideriamo le questioni con ordine.

Il testo

Il *testo oggetto di comprensione* deve

- essere dimensionato sulla competenza dei candidati (non essere troppo facile né troppo difficile);
- risultare interessante a chi lo legge/ascolta;
- rispettare le proprietà di un testo (*in primis*, coerenza e coesione);
- rispecchiare le abitudini di lettura/ascolto dei candidati;
- essere sufficientemente informativo affinché vi si possano ricavare degli *item*;
- trattare di temi familiari ai candidati;
- evitare di toccare argomenti sensibili, che possono suscitare disagio nell'ascoltatore/lettore (guerra, malattie, morte, ecc.).

Per gli audio, in particolare, occorre accertarsi che

- la qualità della registrazione sia eccellente;
- la velocità di eloquio non costituisca un ostacolo alla comprensione;
- tratti regionali o espressioni gergali non inficino la comprensione.

Le attività

Così come il testo deve rispecchiare le abitudini di lettura/ascolto di un candidato, alla pari le *attività* scelte dal valutatore devono risultare familiari a coloro ai quali sono destinate. In caso contrario, la non dimestichezza con la tecnica può esercitare un effetto negativo sulla prestazione. Tra le attività di ascolto/lettura più note, menzioniamo:

- Il *vero/falso*. Facilissima da realizzare, questa tecnica ammette tuttavia una certa arbitrarietà nella risposta: l'apprendente ha il 50% di probabilità di indovinare l'opzione esatta. Per ovviare al problema, in genere si adottano due soluzioni:
 - a) elaborare un numero considerevole di *item* (l'esercizio è noto con il termine *individuazione di informazioni*; è presente, per esempio, nella certificazione di italiano CILS);
 - b) richiedere allo studente di motivare la risposta (*vero, perché? falso, perché?*), contenendo il numero di parole di cui la stessa deve essere composta (per esempio nella certificazione di lingua inglese IELTS sono ammesse risposte di tre parole al massimo; nell'esame in lingua straniera della maturità in Austria il limite è di quattro). Tale variante è nota con il nome *vero/falso con giustificazione*;

- Il *quesito a scelta multipla*. Si tratta di una tecnica complessa; richiede tempo e una serie considerevole di attenzioni (*in primis*, il fatto che le opzioni siano di pari lunghezza e siano espresse in un linguaggio semplice, breve e immediato; inoltre, i distrattori devono essere plausibili, cioè non assurdi, e quindi non ‘eliminabili’ di per sé). Si applica, in genere, a testi informativamente densi (altrimenti è difficile escogitare opzioni che abbiano un aggancio con il testo);
- Le *domande aperte*. L'apprendente è tenuto a elaborare con parole sue un contenuto veicolato dal testo; si può esprimere nella lingua di origine (posto che il valutatore la comprenda) o nella lingua oggetto di studio (in questo secondo caso, eventuali errori grammaticali o ortografici non vanno penalizzati). Al fine di un maggior controllo delle risposte, è possibile imporre un numero massimo di parole (come per il *vero/falso con giustificazione*, di cui abbiamo parlato sopra);
- L'*abbinamento*. Prevede la combinazione di elementi distinti (es. immagini-enunciati) riferiti al testo. Generalmente sono presenti degli intrusi (es. due o tre immagini non abbinabili), la presenza dei quali vale a impedire che si diano combinazioni automatiche (per esempio, se avessimo un pari numero di immagini e di enunciati, una volta eseguita la penultima combinazione della serie, il candidato saprebbe automaticamente qual è l'ultima);
- Il *riordino*. È una tecnica impiegata solo per la valutazione della comprensione scritta. Prevede uno spezzettamento del testo: il candidato è tenuto a ricomporre la sequenza originaria. La tecnica si accompagna a testi lineari, quali una biografia o una narrazione semplice. Occorre, in ogni caso, far attenzione affinché lo

spezzettamento non sia eccessivo, onde evitare di imporre un impegno cognitivo eccessivo a carico del lettore;

- L'*incastro*. È una tecnica impiegata solo per la valutazione della comprensione scritta. Il candidato è tenuto ad inserire, all'interno del testo, parti espunte, che gli vengono presentate in disordine. Vale la pena, anche in questo caso, prevedere degli intrusi.

Vi sono, in aggiunta, tecniche di comprensione scritta verso le quali esprimiamo una certa cautela. Il *completamento di tabella*, per esempio, può non risultare familiare a molti candidati. Il *cloze*, d'altro canto, è un'attività che induce a una comprensione 'a macchie di leopardo': lo studente completa le parole espunte appoggiandosi al co-testo (le parole che vengono prima e dopo lo spazio bianco), senza essere stimolato a un'interpretazione generale dei contenuti (TORRESAN, 2015). In riferimento alla comprensione audio, altrove dichiarammo invalido il *dettato*, posto che non è chiaro cosa valuti effettivamente (TORRESAN, 2014; cfr. anche BALBONI, 1998, 2008.).

Gli item

Oltre al *testo* e al tipo di *attività* da impiegare, al valutatore è richiesto di operare una redazione accurata degli *item*. Possiamo concepire la confezione di un test come un vero e proprio genere di scrittura: la competenza di un *item writer* si sviluppa nel tempo, mediante un esercizio ripetuto, proprio come avviene con la competenza di un qualsiasi scrittore. L'attenzione dell'*item writer* va diretta su più direzioni: il singolo *item* deve risultare indipendente dagli *item* che compongono la stessa batteria (cioè non deve riferirsi al medesimo nucleo informativo su cui insiste un altro *item*) e, al tempo stesso, dev'essere

omogeneo con gli altri *item* (è necessario abbia per oggetto lo stesso costrutto)¹; deve essere calibrato in termini cognitivi (prevedendo l'attivazione di strategie di decodificazione/interpretazione che si rapportano al livello di competenza del candidato) e, al tempo stesso, dev'essere linguisticamente accessibile (va formulato in maniera chiara e semplice, senza comportare difficoltà di interpretazione), e così via. Chi voglia approfondire la questione, può consultare le raccomandazioni da noi redatte, accessibili online sia in italiano (TORRESAN, 2019) che in portoghese (TORRESAN, 2020a).

¹ In parole povere, se il costrutto è la lettura, l'*item* non può sollecitare solo conoscenze lessicali ma deve comportare un'interpretazione dei significati.

Avaliar a compreensão

Ao estimar as habilidades de compreensão de um indivíduo, a atenção do avaliador deve estar em vários planos:

- *o texto a ser compreendido;*
- *o tipo de atividade;*
- *os elementos (itens) que compõem a própria atividade.*

Vejamos as questões mais detalhadamente.

O texto

O texto a ser compreendido deve

- ser adequado à competência dos candidatos (nem fácil demais e nem muito difícil);
- despertar o interesse do leitor/ouvinte;
- respeitar as propriedades de um texto (em primeiro lugar, coerência e coesão);
- refletir os hábitos de leitura/escuta dos candidatos;
- ser suficientemente informativo, permitindo que os itens sejam retirados do próprio texto;
- tratar de assuntos com os quais os candidatos tenham familiaridade;
- evitar tocar em temas sensíveis, que podem causar desconforto no ouvinte/leitor (guerra, doença, morte, etc.).

Em especial no que se refere ao áudio, é necessário garantir que

- a qualidade da gravação seja excelente;
- a velocidade da fala não constitua um obstáculo à compreensão;
- características regionais ou gírias não afetem a compreensão.

As atividades

Assim como o texto deve refletir os hábitos de leitura/escuta de um candidato, as *atividades* escolhidas pelo avaliador também devem ser familiares àqueles para os quais se destinam. Caso contrário, a falta de familiaridade com a técnica pode ter um efeito negativo sobre o desempenho. Citamos, a seguir, as atividades de escuta/leitura mais populares:

- *Verdadeiro/falso*. Muito fácil de executar, esta técnica admite, no entanto, uma certa arbitrariedade na resposta: o aluno tem 50% de chance de adivinhar a opção correta. Para superar o problema, duas soluções são geralmente adotadas:
 - a) elaborar um número considerável de itens (o exercício é conhecido pelo termo *identificação de informações*; está presente, por exemplo, no exame de proficiência de italiano CILS);
 - b) pedir para o estudante justificar a resposta (*verdadeiro, por quê? falso, por quê?*), especificando o número de palavras permitidas nas respostas (por exemplo, no exame de proficiência de língua inglesa IELTS, as respostas devem ter no máximo três palavras; já no exame de língua estrangeira para obtenção do diploma de Ensino Médio na Áustria, o

limite é de quatro palavras). Tal técnica é chamada de *verdadeiro/falso com justificativa*;

- *Questão de múltipla escolha*. Esta é uma técnica complexa; requer tempo e uma série de atenções (em primeiro lugar, as opções devem ter o mesmo tamanho e devem ser escritas com uma linguagem simples, curta e imediata; além disso, os distratores devem ser plausíveis, ou seja, com sentido, e, portanto, não ‘elimináveis’ automaticamente). Utilizam-se, geralmente, textos repletos de informações (caso contrário, é difícil inventar opções que tenham uma ligação com o texto);
- *Perguntas abertas*. O aluno deve elaborar com suas próprias palavras um conteúdo transmitido pelo texto, que pode ser expresso na língua de origem (desde que o avaliador compreenda) ou na língua de estudo (neste último caso, erros gramaticais ou ortográficos não devem ser penalizados). A fim de melhor controlar as respostas, é possível estabelecer um número máximo de palavras (como para o *verdadeiro/falso com justificativa*, mencionado acima);
- *Questões de correspondência*. Os alunos correlacionam itens distintos (ex. imagens-enunciados) referentes ao texto. Normalmente aparecem intrusos (ex. duas ou três imagens que não podem ser combinadas), cuja presença consiste em impedir combinações automáticas (por exemplo, se tivéssemos um número igual de imagens e enunciados, o candidato, depois de ter feito a penúltima combinação da série, saberia automaticamente qual é a última combinação);
- *Organização do texto*. É uma técnica utilizada apenas para a avaliação da compreensão escrita. O candidato, partindo de um texto fragmentado, deve reconstruir a sequência original. A técnica é

indicada para textos lineares, como uma biografia ou uma narrativa simples. Em todo caso, é necessário prestar atenção para que a fragmentação não seja excessiva, de modo a não exigir um excesso de esforço cognitivo por parte do leitor;

- *Preenchimento de lacunas*. É uma técnica utilizada apenas para a avaliação da compreensão escrita. O candidato deve completar o texto com as partes que faltam. Estas partes são apresentadas fora de ordem e, neste caso, também vale a pena inserir intrusos.

Existem, além disso, técnicas de compreensão escrita que exigem uma certa cautela. O *preenchimento de tabela*, por exemplo, pode não ser familiar para muitos candidatos. O *cloze*, por outro lado, é uma atividade que leva a uma compreensão através de ‘pistas’: o estudante completa com as palavras que faltam tomando por base o contexto (as palavras que vêm antes ou depois do espaço em branco), sem ser estimulado a fazer uma interpretação geral dos conteúdos (TORRESAN, 2015). Com relação à compreensão de áudio, já declaramos o *ditado* inválido, uma vez que não fica claro o que realmente está sendo avaliado (TORRESAN, 2014; cfr. também BALBONI, 1998, 2008.).

Os itens

Além do *texto* e do tipo de *atividade* a ser utilizada, é necessário que o avaliador redija os itens com exatidão. Podemos conceber a elaboração de um teste como um verdadeiro gênero textual: a competência de um redator de itens se desenvolve ao longo do tempo, através da prática, assim como acontece com a competência de qualquer escritor. A atenção do redator de itens aponta para diversas direções: um item deve ser independente dos outros itens que compõem o banco de itens (ou seja, não deve fazer referência ao núcleo informativo ao qual se refere outro item) e, ao mesmo tempo, deve ser

coerente com os outros itens (é necessário que tenha o mesmo constructo)²; deve ser adequado em termos cognitivos (prevendo a ativação de estratégias de decodificação/interpretação que se relacionem com o nível de competência do candidato) e, ao mesmo tempo, deve ser linguisticamente acessível (deve ser formulado de forma clara e simples, sem causar dificuldades de interpretação), e assim por diante. Quem quiser saber mais sobre isso, pode consultar nossas recomendações disponíveis on-line, tanto em italiano (TORRESAN, 2019) quanto em português (TORRESAN, 2020a).

² Simplificando, se o constructo é a leitura, o *item* não pode demandar apenas conhecimentos lexicais, mas deve abranger uma interpretação dos significados.

Valutare la produzione

Anche la valutazione della produzione linguistica prevede un'attenzione a più livelli:

- *lo stimolo (prompt) che induce l'allievo a parlare/scrivere;*
- *la scala in adozione;*
- *il giudizio del valutatore.*

Illustriamo i tre aspetti con ordine.

Lo stimolo

Lo *stimolo* (o *prompt*) è rappresentato dalle indicazioni mediante le quali l'apprendente è tenuto a sviluppare un certo tipo di testo o un certo tipo di discorso. Alcune regole di base vanno rispettate:

- la formulazione dev'essere immediata e breve;
- se ci sono immagini di accompagnamento, è bene che siano chiare (non ambigue né connotate culturalmente);
- le operazioni linguistico-cognitive richieste devono rispecchiare quelle di cui dà prova uno scrittore o un parlante in un contesto di uso spontaneo della lingua;
- il compito dev'essere dimensionato in rapporto al livello di competenza dello scrittore/parlante;
- le coordinate sociocomunicative devono essere, per quanto possibile, esplicite (allo scrittore/parlante dev'essere chiaro qual è il suo ruolo,

chi è il destinatario del messaggio, qual è il proposito comunicativo, qual è il genere);

- è necessario elicitarne un campione relativamente esteso di lingua;
- è necessario andare incontro agli interessi dell'apprendente (a tal proposito, a questi può essere concesso di scegliere un'opzione all'interno di un ventaglio di possibilità);
- al candidato può essere concesso del tempo affinché organizzi i pensieri;
- per lo scritto in particolare, è bene prevedere del tempo nel quale l'apprendente riveda l'elaborato;
- per il parlato in particolare, al fine di arginare uno stato d'ansia, il valutatore/esaminatore deve assicurare che il compito venga svolto in un'atmosfera serena e rilassata;
- sempre in riferimento al parlato, si può pensare a un doppio compito:
 - o *monologo* (la presentazione di un certo tema);
 - o *interazione* (sotto forma di intervista, *role-play* o dialogo a tema tra l'apprendente e un compagno o tra l'apprendente e un interlocutore esterno).³

³ Nel caso dell'*interazione orale apprendente-interlocutore* esterno, per garantire una maggiore affidabilità al giudizio, è bene che chi valuta non sia anche colui che, al tempo stesso, deve interagire con il candidato. In altre parole, è bene che il valutatore valuti la *performance* senza prendere parte all'interazione con il candidato; una terza persona (un collega o un'assistente) veste i panni dell'intervistatore.

Per quanto concerne l'*interazione tra compagni* va rilevato il fatto che se, da un lato, essa si presenta più equilibrata in termini di ruoli (i due sono alla pari) e genera una negoziazione di significati maggiore rispetto a quella che si dà nell'interazione con un interlocutore esterno, dall'altro richiede cautela nella formazione delle coppie. Una grande differenza di competenza tra i candidati rischia di inibire, anziché facilitare, la comunicazione (LOUMA, 2004).

La scala di valutazione

Al fine di valutare una *performance* linguistica il valutatore può servirsi di una scala semplice, composta di intervalli numerici (es. 0-10), oppure definita in forma di tabella e corredata di descrittori, i quali possono riferirsi alternativamente

- a una competenza in generale (riassumono, ciascuno, più componenti; ad esempio, per lo scritto: ortografia, grammatica, lessico, coesione e coerenza, ecc.); è il caso della *griglia olistica*;
- a componenti separate; è il caso della *griglia analitica* (della quale abbiamo riportato un esempio all'appendice 1).

La *griglia olistica*, per quanto agile da usare, è più approssimativa. Quella *analitica*, di contro, benché più laboriosa, consente una rappresentazione molto dettagliata della competenza e costituisce la base di un *feedback* mirato: il valutato viene informato dell'aspetto specifico da curare.

Le categorie di cui si compone una *griglia analitica* possono avere un egual valore l'una rispetto all'altra oppure assumere un peso diverso. La diversità, in tal caso, è dettata dall'importanza che la componente assume rispetto alle altre, in considerazione al livello di competenza e/o al profilo dei candidati. Per esempio, nella valutazione dello scritto di un apprendente di livello A1 o A2⁴, alla *componente grammaticale* può essere attribuito un valore minore rispetto a quello riconosciuto all'*efficacia*, posto che ai livelli bassi è più importante farsi capire, rispetto al fatto di esprimersi in maniera accurata.

⁴ Cfr. *Quadro Comune Europeo di Riferimento per le Lingue* (COUNCIL OF EUROPE, 2001).

Il giudizio del valutatore

Esiste una convinzione ingenua, in virtù della quale si ritiene che il voto assegnato a una composizione scritta o a un campione di lingua orale ci restituisca esattamente il livello di competenza e che il valutatore agisca come un semplice testimone. I dati che provengono dalla ricerca ci dicono che così non è: il giudizio attribuito da due valutatori a uno stesso elaborato può oscillare dal 30% al 60% (cfr. McNAMARA, 1996: 122). Ciò significa che, con un'oscillazione del 30%, a uno stesso candidato, su una scala da 0 a 10, potrebbe essere assegnato il voto 4,5 dal valutatore X o il voto 7.5 dal valutatore Y.

La variabilità può essere legata a comportamenti stabili o a circostanze aleatorie. E cioè il valutatore può essere

- *sempre* più severo/generoso rispetto a un altro (e ciò può verificarsi in riferimento a tutte le componenti che concorrono al costrutto oppure a una in particolare);
- ora severo ora meno severo, sulla base del grado di stanchezza o dell'umore del momento, di questioni di simpatia/antipatia verso il candidato, dell'apprezzamento per la calligrafia; ecc.

Come ridurre i margini di variabilità, al fine di garantire un giudizio per quanto possibile normato? La soluzione ottimale pare essere quella di allestire occasioni di confronto tra valutatori. In queste circostanze, ciascun valutatore può condividere con gli altri il voto assegnato e giustificarlo; laddove ci sono divergenze, le argomentazioni dell'uno possono convincere l'altro a convenire su un unico punteggio. Le esperienze finora condotte ci informano che, grazie al confronto tra le valutazioni (cfr. McNAMARA, 1996),

- si mettono a nudo quei fattori idiosincratici che determinano una erraticità del giudizio (il fatto cioè che lo stesso esaminatore non sia coerente con sè stesso);
- si contiene la variazione di severità/generosità entro certi parametri, pur non potendo eliminarla del tutto.

Avaliar a produção

A avaliação da produção linguística também requer atenção em vários níveis:

- *o estímulo (prompt) que leva o aluno a falar/ escrever;*
- *a escala escolhida;*
- *a avaliação do avaliador.*

A seguir, apresentamos os três aspectos detalhadamente.

O estímulo

O *estímulo* (ou *prompt*) é representado pelas indicações com as quais o aluno deve desenvolver um certo tipo de texto ou um certo tipo de discurso. Algumas regras básicas devem ser respeitadas:

- a formulação deve ser breve e sucinta;
- se houver imagens de apoio, estas devem ser claras (não devem ser ambíguas ou trazer conotações culturais);
- as operações linguístico-cognitivas necessárias devem refletir aquelas demonstradas por um escritor ou um falante em um contexto de uso espontâneo da língua;
- a tarefa deve estar adequada ao nível de competência do escritor/falante;
- as coordenadas sociocomunicativas devem ser tão explícitas quanto possível (é preciso que fique claro para o escritor/falante qual é o seu papel, quem é o destinatário da mensagem, qual é o propósito comunicativo, qual é o gênero);

- é necessária uma amostra relativamente ampla de língua;
- é necessário atender aos interesses dos alunos (a este respeito, pode-se permitir aos estudantes que escolham uma opção dentro de um leque de possibilidades);
- o candidato pode dispor de tempo para organizar os seus pensamentos;
- é aconselhável, especialmente para a escrita, prever o tempo durante o qual o aluno revê o trabalho;
- especialmente no que se refere à fala, a fim de conter a ansiedade, o avaliador/examinador deve garantir que a tarefa seja realizada em um ambiente sereno e descontraído;
- sempre com referência à fala, você pode pensar em uma dupla tarefa:
 - o *monólogo* (apresentação de um determinado tema);
 - o *interação* (sob a forma de entrevista, *role-play* ou diálogo temático entre o estudante e um colega ou entre o estudante e um interlocutor externo).⁵

A escala de avaliação

A fim de avaliar uma *performance* linguística, o avaliador pode usar uma escala simples, constituída por intervalos numéricos (ex. 0-10), ou uma escala

⁵ No caso da *interação oral estudante-interlocutor* externo, para garantir uma maior confiabilidade à avaliação, é aconselhável que quem avalia não seja também aquele que, ao mesmo tempo, deve interagir com o candidato. Em outras palavras, é aconselhável que o avaliador avalie a *performance* sem participar da interação com o candidato; uma terceira pessoa (um colega ou um assistente) assume o papel do entrevistador.

No que diz respeito à *interação entre estudantes*, nota-se que, por um lado, é mais equilibrada em termos de papéis (os dois estão de igual para igual) e gera uma negociação mais significativa em relação àquela que ocorre na interação com um interlocutor externo, por outro lado, exige cautela na formação das duplas. Uma grande diferença de competência entre os candidatos corre o risco de inibir, em vez de facilitar, a comunicação (LOUMA, 2004).

definida em forma de tabela e acompanhada de descritores, que podem referir-se alternadamente

- a uma competência em geral (resumindo, em cada uma, vários componentes; por exemplo, para a escrita: ortografia, gramática, léxico, coesão e coerência, etc.); este é o caso da *tabela holística*;
- a componentes separados; este é o caso da *tabela analítica* (cujo exemplo consta no [apêndice 1](#)).

A tabela *holística*, ainda que seja flexível, é mais aproximativa. A *analítica*, por outro lado, embora mais trabalhosa, permite uma representação muito detalhada da competência e constitui a base de um *feedback* direcionado: o candidato é informado sobre um aspecto específico a ser melhorado.

As categorias que compõem uma *tabela analítica* podem ter um valor igual ou assumir um peso diferente. A diversidade, neste caso, é ditada pela importância que o componente assume em relação aos demais, levando em conta o nível de competência e/ou o perfil dos candidatos. Por exemplo, na avaliação do nível de escrita de um aluno A1 ou A2⁶, pode-se atribuir ao *componente gramatical* um valor menor em comparação ao valor conferido à *eficácia*, uma vez que nos níveis mais baixos é mais importante ser compreendido do que saber se expressar de maneira correta.

A avaliação do avaliador

Há uma crença ingênua de que a nota dada a uma produção escrita ou a uma produção oral nos revela exatamente o nível de competência e que o avaliador age como uma mera testemunha. Os dados da pesquisa nos dizem que este não é o caso: a avaliação feita por dois avaliadores para uma mesma

⁶ Cfr. *Quadro Comum Europeu de Referência para as Línguas* (COUNCIL OF EUROPE, 2001).

tarefa pode variar de 30% a 60% (cfr. McNAMARA, 1996: 122). Isso significa que, com uma variação de 30%, o mesmo candidato, em uma escala de 0 a 10, poderia receber a nota 4,5 do avaliador X ou a nota 7.5 do avaliador Y.

A variabilidade pode estar ligada a comportamentos estáveis ou a circunstâncias aleatórias. O avaliador pode ser

- *sempre* mais severo/generoso em relação a outro avaliador (e isso pode ocorrer com relação a todos os componentes que contribuem para o constructo ou a um componente em particular);
- às vezes severo, às vezes menos severo, dependendo do nível de cansaço ou do estado de humor daquele momento, de questões de simpatia/antipatia com o candidato, de apreço pela caligrafia; etc.

Como reduzir as margens de variabilidade, a fim de garantir uma avaliação padronizada sempre que possível? A solução ideal parece ser criar oportunidades de comparação entre avaliadores. Nestas circunstâncias, cada avaliador pode partilhar e justificar a nota atribuída com os outros; onde há divergências, os argumentos de um podem convencer o outro a concordar com uma única pontuação. As experiências até agora demonstram que, graças à comparação entre as avaliações (cfr. McNAMARA, 1996),

- vêm à tona os fatores idiossincráticos que determinam uma instabilidade na avaliação (isto é, o fato de que o examinador não é coerente consigo mesmo);
- a variação de severidade/generosidade se reduz a determinados parâmetros, embora ela não possa ser completamente eliminada.

Valutare la mediazione

La *mediazione linguistica* indica un'azione condotta su più testi. Si parla, più in particolare, di

- *mediazione intralinguistica*, qualora i testi oggetto di mediazione siano veicolati in una stessa lingua (esempi di attività sono il *riassunto*, la *variazione di registro* e la *trasposizione di genere*);
- *mediazione interlinguistica*, quando invece i testi oggetto di mediazione sono veicolati in due lingue diverse (è il caso della *traduzione*).

In entrambe le circostanze si applica una *valutazione integrata*, che contempla cioè l'integrazione di più abilità linguistiche. Si badi: non si tratta di una semplice somma di abilità ma di un costrutto qualitativamente diverso. Il *riassunto*, per dire, non è dato dall'addizione della comprensione scritta/orale e della produzione scritta. È un qualcosa di più: comporta, in effetti, delle operazioni cognitive che gli sono proprie, quali la selezione e la riorganizzazione dei concetti. Ciò significa che il *riassunto* (così come ogni altra attività di *mediazione*) non può essere adottato né come prova di comprensione né come prova di produzione. Il fatto che abbia uno statuto autonomo richiede l'elaborazione di *standard* specifici, ricavati mediante lo spoglio di *corpora* di attività prodotte dagli allievi. Ciò può portare alla redazione di griglie di valutazione specifiche, simili a quelle utilizzate per valutare le abilità produttive⁷. Va in ogni caso riconosciuto che il dibattito inerente la valutazione

⁷ Quali esempi di griglie volte a valutare *abilità integrate*, si vedano quelle in uso dal Trinity College di Londra, volte ad accertare la competenza in inglese: "Speaking and Listening Rating Scale"; "Reading into Writing Rating Scale", <www.trinitycollege.com/qualifications/english-language/ISE/ISE-results-and-certificates/ISE-rating-scales>.

delle *abilità di mediazione* è relativamente recente; la letteratura di riferimento si sta lentamente consolidando (MAGRIS, 2006; HOUSE, 2013; WILLIAMS, 2013; CUMMING, 2014).

Avaliar a mediação

A *mediação linguística* indica ações tomadas em vários textos. Mais especificamente, fala-se de

- *mediação intralinguística*, em que os textos mediados são veiculados na mesma língua (exemplos de atividades são *resumo*, *alteração de registro* e *alteração de gênero textual*);
- *mediação interlinguística*, quando os textos mediados são veiculados em duas línguas diferentes (é o caso da *tradução*).

Em ambos os casos é aplicada uma *avaliação integrada*, que abrange a integração de várias habilidades linguísticas. Lembre-se: não se trata de uma simples soma de habilidades mas de um constructo qualitativamente diferente. O *resumo*, por exemplo, não se dá através da soma de compreensão escrita/oral e produção escrita. É algo a mais: envolve, na verdade, operações cognitivas próprias, como a seleção e a reorganização de conceitos. Isto significa que o *resumo* (bem como qualquer outra atividade de *mediação*) não pode ser adotado nem como prova de compreensão nem como prova de produção. O fato de o resumo ter estatuto autônomo requer a elaboração de *padrões* específicos, que são obtidos através da análise de uma série de atividades produzidas pelos alunos. Isto pode levar à elaboração de tabelas específicas de avaliação, semelhantes às tabelas utilizadas para avaliar as habilidades produtivas⁸. De qualquer forma, deve-se reconhecer que o debate em torno da avaliação das

⁸ Exemplos de tabelas para avaliar *habilidades integradas* são aquelas usadas pelo Trinity College em Londres para verificar a competência em inglês: “Speaking and Listening Rating Scale”; “Reading into Writing Rating Scale”, <www.trinitycollege.com/qualifications/english-language/ISE/ISE-results-and-certificates/ISE-rating-scales>.

habilidades de mediação é relativamente recente; a literatura de referência vem se consolidando lentamente (MAGRIS, 2006; HOUSE, 2013; WILLIAMS, 2013; CUMMING, 2014).

Conclusioni

Si ci chiedessimo, in una nota finale, quali sono i fattori che concorrono alla qualità del giudizio, segnaleremmo

- *il pensiero critico;*
- *un volume significativo di dati;*
- *il lavoro di squadra.*

Approfondiamoli.

Pensiero critico

Il buon valutatore, dicevamo, sa che non è possibile attingere a una verità oggettiva circa la competenza di una persona; è anzi convinto che ogni forma di valutazione delle facoltà umane restituisce una fotografia approssimativa della realtà (CERVINI, MARTARI, 2020). Nell'impresa di limitare, per quanto concesso, tali margini di approssimazione, egli, durante l'allestimento di una prova, è tenuto a considerare una serie ampia di variabili, triangolando osservazioni e attenzioni di vario genere. È un ragionamento reticolare che abbraccia numerosissimi elementi. In altre parole, valutare non è un'operazione lineare, progressiva, ma iterativa, complessa, sfaccettata.

Un volume significativo di dati

Capita di sentir dire che la valutazione condotta su larga scala è valida e affidabile, mentre quella svolta dagli insegnanti in classe è soggettiva, potenzialmente viziata da pregiudizi e favoritismi. Si tratta di una convinzione discutibile (ALLAN, 1999). Se, in effetti, da un lato, la scarsa formazione e il fatto

che l'insegnante si trovi ad operare da solo, senza la possibilità di confrontarsi con i colleghi, possono minare l'attendibilità del giudizio, dall'altro, egli conta su un'osservazione più *estesa* rispetto a quella realizzabile in un contesto certificatorio. Va da sé che un'osservazione prolungata nel tempo del comportamento del singolo consente di formulare un *feedback* più attendibile.

Per converso, allo scopo di raggiungere una certa affidabilità, una qualsiasi prova standardizzata, applicata su larga scala, deve poggiare su attività variegata per tipologia, in modo da coprire diversi aspetti del costrutto di riferimento (cfr. ONAIBA, JANAT, 2019), ed essere applicata a un numero ampio di candidati, in maniera tale da disporre di evidenze sufficienti affinché se ne possano monitorare gli esiti.

Il lavoro di squadra

Nel 1906 a Plymouth, nel sud dell'Inghilterra, in una fiera veniva lanciata una sfida: indovinare il peso di un bue. Francis Galton, psicologo ed esperto di statistica, raccolse i pareri dei 787 partecipanti alla gara. Notò che la mediana delle ipotesi (il valore collocato perfettamente a metà nella serie di dati) variava dell'1 per mille rispetto al peso reale dell'animale. Uno scarto irrisorio! I 787 partecipanti presi nel loro insieme avevano azzeccato, con una precisione sorprendente, il peso dell'animale (GALTON, 1907).

Esperimento semplice, dalle conclusioni immediate: per affinare il giudizio il confronto tra più pareri è fondamentale. La collegialità, ovvero il fatto di cooperare con altri durante la realizzazione e il controllo di una prova, è una garanzia di qualità, posto che ciò che sfugge all'uno può essere notato dall'altro. Il pensiero di Paulo Freire si rivela attuale anche nell'ambito dell'elaborazione di un test (1987): "*Ninguém ignora tudo, ninguém sabe tudo, todos nós sabemos alguma coisa. Todos nós ignoramos alguma coisa, por isso aprendemos sempre*".

Conclusões

Ao nos perguntarmos, em observação final, quais são os fatores que contribuem para a qualidade da avaliação, assinalamos

- *pensamento crítico;*
- *um volume significativo de dados;*
- *trabalho em equipe.*

A seguir, uma análise de cada fator.

Pensamento crítico

O bom avaliador, como já dissemos, sabe que não é possível obter uma verdade objetiva a respeito da competência de uma pessoa; pelo contrário, está convencido de que toda forma de avaliação das faculdades humanas dá uma imagem aproximada da realidade (CERVINI, MARTARI, 2020). Com o intuito de limitar, na medida do permitido, essas margens de aproximação, o avaliador, durante a preparação de uma prova, deve considerar uma ampla série de variáveis, articulando observações e atenções de diversos tipos. É um raciocínio complexo que abraça inúmeros elementos. Em outras palavras, avaliar não é uma operação linear, progressiva, mas sim iterativa, complexa, multifacetada.

Um volume significativo de dados

Ouve-se que a avaliação em larga escala é válida e confiável, ao passo que a avaliação realizada pelos professores em sala de aula é subjetiva, potencialmente corrompida por preconceitos e favoritismos. Trata-se de uma convicção questionável (ALLAN, 1999). Se, por um lado, a falta de formação

e o fato de o professor trabalhar sozinho, sem a possibilidade de se confrontar com os colegas, podem minar a credibilidade da avaliação, por outro lado, ele conta com uma observação mais *ampla* do que aquela que pode ser feita em um contexto de proficiência. É óbvio que uma observação a longo prazo do comportamento do indivíduo permite emitir um *feedback* mais confiável.

Em contrapartida, para alcançar uma certa confiabilidade, qualquer prova padronizada, aplicada em larga escala, deve contar com atividades variadas por tipo, de modo a atender diversos aspectos do constructo de referência (cfr. ONAIBA, JANAT, 2019), e deve ser aplicada a um grande número de candidatos, de modo que existam evidências suficientes para monitorar os resultados dos estudantes.

Trabalho em equipe

Em 1906, em Plymouth, no sul da Inglaterra, um desafio foi lançado em uma feira: adivinhar o peso de um boi. Francis Galton, psicólogo e perito em estatística, coletou as opiniões dos 787 participantes. Ele observou que a mediana das hipóteses (o valor perfeitamente colocado no meio da série de dados) variou em 1 por mil, em relação ao peso real do animal. Um resíduo irrisório! Os 787 participantes, tomados como um todo, acertaram, com surpreendente precisão, o peso do animal (GALTON, 1907).

Experimento simples, com conclusões imediatas: para refinar a avaliação, a comparação entre várias opiniões é fundamental. A colegialidade, ou seja o fato de cooperar com os outros durante a realização e o controle de uma prova, é uma garantia de qualidade, já que o que escapa de um pode ser notado pelo outro. O pensamento de Paulo Freire também se revela atual no contexto da elaboração de um teste (1987): “*Ninguém ignora tudo, ninguém sabe tudo, todos nós sabemos alguma coisa. Todos nós ignoramos alguma coisa, por isso aprendemos sempre*”.

Parte II
Glossario

Parte II
Glossário

Introduzione

Lo scopo di un glossario è definire termini specifici di un determinato ambito. Il presente glossario fornisce una mappa dei termini ricorrenti nella valutazione linguistica.

Le voci (circa duecento) contemplano sia aspetti generali che aspetti specifici, scendendo nel dettaglio di misurazioni e valori, gran parte dei quali coincidono con quelli della ricerca quantitativa.

Riportate in ordine alfabetico in riferimento all'italiano, le voci sono suddivise nei seguenti ambiti:

- A. Concetti generali
- B. Elementi
- C. Tipologie
 - Tipi di valutazione
 - Tipi di esami
 - Tipi di esercitazioni
- D. Proprietà
 - Proprietà una prova, di un'attività e di una prestazione
 - Proprietà di *prompt*, istruzioni e *item*
- E. Operazioni
- F. Istituzioni e individui
- G. Criticità
- H. Valori, rappresentazioni e misurazioni
- I. Metavalutazione

Nelle definizioni sono presenti le seguenti evidenziazioni.

	VERSIONE IN ITALIANO	VERSIONE IN PORTOGHESE
<i>Parole o espressioni che rimandano a voci generali</i>	in neretto	in azzurro chiaro marcato
<i>Parole o espressioni che rimandano a strumenti di ricerca e a parametri di indagine</i>	in rosso marcato	in azzurro scuro marcato

Introdução

O objetivo de um glossário é definir termos específicos de um determinado campo. Este glossário fornece um mapa dos termos utilizados na avaliação linguística.

Os termos (cerca de duzentos) cobrem aspectos gerais e aspectos específicos, analisando detalhadamente medições e valores. Estes valores coincidem com os valores da pesquisa quantitativa.

Os termos estão na ordem alfabética do italiano e subdividem-se nos seguintes grupos:

- A. Conceitos gerais
- B. Elementos
- C. Tipologias
 - Tipos de avaliação
 - Tipos de exames
 - Tipos de atividades
- D. Propriedades
 - Propriedades de uma prova, de uma atividade
 - Propriedades de *prompt*, instruções, *item*
- E. Procedimentos operacionais
- F. Pessoas e instituições
- G. Aspectos críticos
- H. Medições, representações e valores
- I. Meta-avaliação

Nas definições utilizamos os seguintes destaques.

	VERSÃO EM ITALIANO	VERSÃO EM PORTUGUÊS
<i>Palavras ou expressões referentes a termos gerais</i>	em negrito	cor da fonte: azul claro, em negrito
<i>Palavras ou expressões referentes a ferramentas de pesquisa e a parâmetros quantitativos</i>	cor da fonte: vermelho, em negrito	cor da fonte: azul escuro, em negrito

A. CONCETTI GENERALI

A. CONCEITOS GERAIS

Illustriamo concetti trasversali a qualsiasi tipo di valutazione.

Apresentamos conceitos transversais a todo tipo de avaliação.

Abilità linguistica

Forma di uso della lingua orientato ai significati.

Le *abilità linguistiche* possono essere distinte in

- *primarie* (leggere, ascoltare, parlare, scrivere)
- *integrate* (legate alla combinazione di attività primarie; è il caso del riassunto, della traduzione, ecc.)

Le *abilità primarie*, a loro volta, possono essere distinte in

- *ricettive* o *di comprensione* (leggere, ascoltare);
- *produttive* o *di produzione* (parlare, scrivere).

L'*interazione orale/ scritta* può essere intesa come una *abilità integrata* (poiché integrazione di *abilità ricettive* e *produttive*).

All'interno dei **Documenti Europei** – *Quadro* (COUNCIL OF EUROPE, 2001) e *Companion* (COUNCIL OF EUROPE, 2020) – le *abilità linguistiche* vengono distinte in tre gruppi:

- *ricezione e produzione* (abilità primarie);
- *interazione scritta/orale* (forma di *integrazione* tra le abilità primarie);

- *mediazione* (forma di *integrazione* complessa, che riassume i livelli inferiori: *ricezione/produzione* e *interazione*).

L'abilità di *mediazione* rimanda, in pratica, alla capacità di compiere *passaggi interlinguistici* (come nel caso della *traduzione* e dell'*interpretariato*) e/o *intralinguistici* (come avviene nella manipolazione di testi a vari livelli: *riassunto*, *trasformazione di registro*, *trasposizione di modalità* [dal parlato allo scritto e viceversa], ecc.).⁹

Habilidade linguística

Uso da língua voltado para os significados.

As *habilidades linguísticas* podem ser divididas em

- *primárias* (ler, ouvir, falar, escrever);
- *integradas* (resultado da combinação de atividades primárias; é o que acontece no resumo, na tradução, etc.).

As *habilidades primárias*, por sua vez, podem ser divididas em

- *receptivas* ou *de compreensão* (ler, ouvir);
- *produtivas* ou *de produção* (falar, escrever).

A *interação oral/escrita* pode ser considerada como uma *habilidade integrada* (visto que é integração de *habilidades receptivas* e *produtivas*).

Nos **Documentos Europeus** – *Quadro Comum Europeu* (COUNCIL OF EUROPE, 2001) e *Companion*

⁹ Nel *Companion* (COUNCIL OF EUROPE, 2020) si fa riferimento anche alla *mediazione sociale* (intesa come capacità di creare uno spazio cooperativo e di gestire situazioni conflittuali) e *culturale* (intesa come capacità di creare le condizioni per un confronto pluriculturale).

(COUNCIL OF EUROPE, 2020) – as *habilidades linguísticas* estão divididas em três grupos:

- *recepção e produção* (habilidades primárias);
- *interação escrita/oral* (forma de *integração* das habilidades primárias);
- *mediação* (forma de *integração* complexa, que reúne os níveis inferiores: *recepção/produção* e *interação*).

Na prática, a habilidade de *mediação* se refere à capacidade de realizar *passagens interlinguísticas* (como no caso dos tradutores e intérpretes) e/ou *intra-linguísticas* (como ocorre na reelaboração de textos em vários níveis: *resumo, alteração de registro, transposição de modalidades* [do falado ao escrito e vice-versa], etc.).¹⁰

Argomento di validazione

Raccolta di evidenze a favore del giudizio di **validità** di una prova (cfr. KANE, 1990; 1992; 2001).

Argumento de validação

Coleta de evidências a favor de um parecer de **validade** de uma prova (cfr. KANE, 1990; 1992; 2001).

¹⁰ O *Companion* (COUNCIL OF EUROPE, 2020) também faz referência à *mediação social* (habilidade para criar um espaço cooperativo e lidar com situações de conflito) e *cultural* (habilidade para criar condições que promovam um diálogo pluricultural).

Attività (o esercitazione o tecnica o test format)

Proposta didattica volta a elicitare un comportamento dell'apprendente, soggetto a valutazione.

Più in dettaglio, si parla di *compito* o *esercizio* a seconda che si tratti di un'attività comunicativamente orientata (es. compito di produzione orale) o formalmente orientata (es. esercizio di grammatica).¹¹

Atividade

Proposta didática que visa a avaliar o desempenho do discente.

Mais especificamente, trata-se de *tarefa* ou *exercício*, dependendo se é uma atividade comunicativamente orientada (ex. tarefa de produção oral) ou formalmente orientada (ex. exercício de gramática).¹²

Audit

Controllo dei parametri di una prova ai fini di un accreditamento da parte di un ente riconosciuto (per es. ALTE, *Association of Language Testers in Europe*).

Auditoria

Controle dos parâmetros de uma prova para efeitos de credenciamento por uma instituição reconhecida (p. ex. ALTE, *Association of Language Testers in Europe*).

Ciclo di produzione

Operazioni mediante le quali una prova/un'esercitazione viene elaborata (DOWNING, 2006). Per esempio, nella

¹¹ Per una rassegna delle attività più diffuse nella valutazione linguistica, cfr. Underhill, 1987; Alderson *et al.* 1995; Alderson, 2000; Read, 2000; Buck, 2001; Cushing Weigle, 2002; Purpura, 2004; Louma, 2004; O'Sullivan, 2013; Novello, 2014; Serragiotto, 2016.

¹² Para conferir as atividades mais frequentes na avaliação linguística, cfr. Underhill, 1987; Alderson *et al.* 1995; Alderson, 2000; Read, 2000; Buck, 2001; Cushing Weigle, 2002; Purpura, 2004; Louma, 2004; O'Sullivan, 2013; Novello, 2014; Serragiotto, 2016.

valutazione della comprensione orale/scritta si possono prevedere le seguenti azioni:

- **scelta del *format*;**
- **scelta del testo;**
- **mappatura del testo;**
- **scrittura degli *item*;**
- **revisione;**
- **somministrazione;**
- **assegnazione del punteggio;**
- **comunicazione dei risultati.**

Ciclo de produção

Procedimentos utilizados na elaboração de uma prova/um exercício (DOWNING, 2006). Por exemplo, na avaliação da compreensão oral/escrita é possível prever as seguintes ações:

- **escolha do tipo de prova;**
- **escolha do texto;**
- **mapeamento do texto;**
- **redação dos itens;**
- **revisão;**
- **aplicação;**
- **atribuição da pontuação;**
- **divulgação dos resultados.**

Collegialità

Una prova, un *item* o un'attività, una volta elaborati, costituiscono un materiale 'grezzo' che può essere soggetto a raffinamento. È importante, a tal proposito, che lo sguardo dell'autore si intrecci con gli sguardi di altri, affinché eventuali difetti siano individuati e la

prova, l'*item* o la singola attività abbiano modo di essere perfezionati (cfr. DAVIDSON, LYNCH, 2007).

Colegialidade

Uma prova, um item ou uma atividade, quando elaborados, constituem um material 'bruto' que pode passar por um refinamento. Neste contexto, é importante que o olhar do autor se entrelace com o olhar de outros, permitindo que eventuais defeitos sejam identificados e a prova, o item ou a atividade sejam aperfeiçoados (cfr. DAVIDSON, LYNCH, 2007).

Compromesso

La redazione di un esame linguistico prevede si raggiunga un equilibrio tra diverse proprietà (**economia/praticabilità, validità, impatto, affidabilità**). Una prova massimamente *affidabile* può, in effetti, venire meno in termini di *validità*; così come, al contrario, una prova massimamente *valida* può rischiare di sacrificare una certa *affidabilità*. Nell'allestire una prova è bene, quindi, che si considerino le varie proprietà, e si operino scelte ragionate, in funzione degli obiettivi che ci si prefigge. In prove di certificazione, il valutatore è tenuto a esplicitare nelle **Specifiche** le ragioni di tali scelte.

Compromisso

A elaboração de um exame linguístico deve atingir um equilíbrio entre diversas propriedades (**economia/praticidade, validade, impacto, confiabilidade**). Uma prova extremamente *confiável* pode, na verdade, falhar em termos de *validade*; assim como em sentido inverso, uma prova extremamente

válida pode sacrificar um pouco da *confiabilidade*. Ao construir uma prova, é bom que sejam tomadas em consideração as diversas propriedades, e que sejam feitas escolhas fundamentadas, em função dos objetivos pretendidos. Nas provas de proficiência, o avaliador deve explicar as razões para tais escolhas nas **Especificações**.

Conoscenza linguistica Informazioni in possesso, da parte del candidato, riguardo le proprietà formali della lingua oggetto di studio.

Conhecimento linguístico

Informações que o candidato possui sobre as propriedades formais da língua estudada.

Competenza comunicativa

Capacità dell'allievo di interpretare il significato di testi e di produrre testi.

In un senso stretto il costrutto della competenza comunicativa include **conoscenze** e **strategie**; in un senso lato, comprende pure gli **atteggiamenti** (cfr. PELLERREY, 2004).¹³

Competência comunicativa

Capacidade do estudante de interpretar o sentido dos textos e de produzir textos.

¹³ Per un prospetto su come si è andato evolvendo nel tempo il costrutto della competenza comunicativa, cfr. SCARAMUCCI, 2000; TORRESAN, 2022.

Estritamente falando, o constructo da competência comunicativa inclui **conhecimentos** e **estratégias**; no sentido lato, abrange também as **atitudes** (cfr. PELLERREY, 2004).¹⁴

<i>Competenza specifica</i>	In un senso generale, il concetto si sovrappone a quello di abilità (es. <i>competenza nella lettura, competenza nella scrittura</i>). In un senso stretto, equivale a un uso (esperto) della lingua vincolato a un testo e a un contesto: <i>comprendere testi espositivi; ascoltare annunci; intrattenere un ospite; stendere una relazione; compilare un modulo, ecc.</i>
<i>Competência específica</i>	De modo geral, este conceito se sobrepõe ao de habilidade (ex. <i>competência leitora, competência escritora</i>). Em sentido estrito, equivale a um uso (especializado) da língua, vinculado a um texto e a um contexto: <i>compreender textos informativos; escutar avisos sonoros; entreter um convidado; escrever um relatório; preencher um formulário, etc.</i>
<i>Componente</i>	Fattore che concorre alla definizione del costrutto (definibile sperimentalmente mediante analisi fattoriale).
<i>Componente</i>	Fator que contribui para a definição do constructo (definível experimentalmente por meio de análise fatorial).

¹⁴ Para um panorama sobre como o constructo da competência comunicativa tem evoluído ao longo do tempo, cfr. SCARAMUCCI, 2000; TORRESAN, 2022.

Costrutto

Idea astratta che dà ragione di un determinato comportamento. Si articola in una serie di **componenti** (o fattori) ed è soggetta a misurazione (FULCHER, DAVIDSON, 2007; BACHMAN, 2007; JAMIESON, 2011).

Possiamo distinguere un *costrutto di riferimento* da un *costrutto operativo*: il primo è, appunto l'idea generale definita dalla ricerca, il secondo rappresenta, invece, l'operativizzazione (ovvero la declinazione concreta) del primo all'interno di un particolare contesto (es. apprendenti/candidati con una certa fascia d'età; con bisogni comunicativi particolari, ecc.). Facciamo un esempio: tutti noi possiamo condividere un'idea generale riguardo all'abilità di lettura, trasversale a diversi tipi di testi scritti (*costrutto di riferimento*); tuttavia, al momento di elaborare un test, si possono tener conto di alcune strategie rispetto ad altre a seconda del profilo dei candidati (in questo senso, per esempio, le strategie di lettura applicate nella selezione di un personale amministrativo possono essere diverse da quelle sollecitate da un test rivolto a studenti che si iscrivono a un corso di lingua per puro piacere).¹⁵

¹⁵ Facciamo un altro esempio, esterno al *testing*. Oggi molti manuali di insegnamento delle lingue straniere si dichiarano *comunicativi*, traendo ispirazione da un certo costrutto di **competenza comunicativa**, in cui alle **abilità linguistiche** è attribuito maggior peso rispetto alla **conoscenza linguistica**. Il modo in cui, però, ciascun manuale operativizza (cioè, applica nel concreto) il costrutto della competenza comunicativa varia: un libro può riservare ampio spazio ad attività di comprensione; un altro può presentare un ricco repertorio di attività di mediazione linguistica, e così via. Ciascun manuale, dunque, declina a proprio modo il *costrutto di riferimento* della **competenza comunicativa**, prefigurandosi certi bisogni/obiettivi da parte dell'utenza.

Constructo

Ideia abstrata que justifica um determinado comportamento. É constituída por uma série de **componentes** (ou fatores) e está sujeita à medição (FULCHER, DAVIDSON, 2007; BACHMAN, 2007; JAMIESON, 2011).

Podemos distinguir um *constructo de referência* de um *constructo operacional*: o primeiro é a ideia geral definida pela pesquisa, o segundo representa a operativização (ou seja, uma adaptação concreta) do primeiro em um contexto específico (ex. alunos/condidatos com uma determinada faixa etária; com necessidades comunicativas especiais, etc.). Vamos dar um exemplo: todos nós podemos compartilhar uma ideia geral a respeito da habilidade de leitura, transversal a diferentes tipos de textos escritos (*constructo de referência*); no entanto, ao preparar um teste, algumas estratégias podem ser levadas em conta em comparação com outras, dependendo do perfil do candidato (nesse sentido, as estratégias de leitura aplicadas na seleção de funcionários administrativos podem ser diferentes das exigidas por um teste destinado a estudantes que se inscrevem em um curso de língua por puro deleite).¹⁶

¹⁶ Vamos dar outro exemplo, fora da avaliação. Hoje em dia, muitos manuais de ensino de línguas estrangeiras se declaram *comunicativos*, cuja inspiração nos remete a um certo constructo de **competência comunicativa**. Tais manuais atribuem maior peso às **habilidades linguísticas** em relação ao **conhecimento linguístico**. No entanto, é variada a forma como cada manual torna operacional (ou seja, aplica na prática) o constructo da competência comunicativa: um livro pode dedicar um espaço amplo para atividades de compreensão; outro pode apresentar um rico repertório de atividades de mediação linguística, e assim por diante. Cada manual, portanto, interpreta à sua maneira o *constructo de referência* da **competência comunicativa**, considerando determinadas necessidades/objetivos por parte do público.

Dominio (o ambito) Contesto sociale di uso della lingua.
I macrodomini definiti dal *Quadro* (COUNCIL OF EUROPE, 2001) e ripresi dal *Companion* (COUNCIL OF EUROPE, 2020) sono quattro: *personale, professionale, pubblico, educativo*.

Domínio Contexto social de uso da língua.
Os macrodomínios definidos pelo *Quadro* (COUNCIL OF EUROPE, 2001) e retomados pelo *Companion* (COUNCIL OF EUROPE, 2020) são quatro: *pessoal, profissional, público, educativo*.

Estrapolazione Estensione dell'interpretazione dei risultati dalla prova alla **conoscenza/competenza** del candidato. In altre parole, il **punteggio** non è solo indicativo dell'abilità del candidato a superare quella determinata prova, ma viene esteso allo svolgimento di compiti che attengano allo stesso **costrutto** e riguardino lo stesso **dominio**/gli stessi domini. Se, per esempio, un candidato risolve con successo compiti di ascolto di livello B1 basati su conversazioni tra amici si potrebbe generalizzare, dicendo che, in un senso lato, egli è in grado di ascoltare (*costrutto*) un vasto repertorio di dialoghi informali (*dominio*) di livello B1 nella lingua oggetto di studio (cfr. OLIVERI *et al.*, 2015: 24-25).

Extrapolação Expansão da interpretação dos resultados obtidos na prova para o **conhecimento/competência** do candidato. Em outras palavras, a **pontuação** indica não só a habilidade do candidato em passar numa prova específica, mas também

demonstra sua capacidade de desempenhar tarefas relacionadas ao mesmo **constructo** e ao mesmo **domínio**/aos mesmos domínios. Por exemplo, se um candidato resolve com sucesso as tarefas de escuta de nível B1, com base em conversas informais, podemos generalizar dizendo que, em um sentido amplo, ele é capaz de ouvir (*constructo*) um vasto repertório de diálogos informais (*domínio*) de nível B1 na língua em estudo (cfr. OLIVERI *et al.*, 2015: 24-25).

Feedback (o retroalimentazione)

Informazione fornita all'esaminando (sottoforma di voto o di commento esteso) in merito a come ha svolto la prova ed eventualmente riguardo a cosa/come può migliorare in futuro.

È bene che il *feedback* sia rispettoso (non va riferito alla persona ma alla prestazione) e circostanziato (non generico).

Feedback

Informação disponibilizada para o candidato (sob a forma de nota ou de comentário por escrito) em relação ao seu desempenho na prova e, eventualmente, sobre o que/como pode melhorar no futuro.

O *feedback* deve ser respeitoso (não deve se referir à pessoa mas ao desempenho) e detalhado (não genérico).

Performance

Prestazione del candidato in prove che hanno per oggetto le abilità produttive o interattive.

Performance Desempenho do candidato em provas com foco nas habilidades produtivas ou interacionais.

Prova Con questo termine si intendono in genere due realtà:

- una singola esercitazione (es. “una prova di comprensione orale”);
- un pacchetto di attività (es. Prova TOEFL o Prova CELPE-Bras).

In questa sede, con il termine “prova” ci riferiamo alla seconda accezione.¹⁷

Prova Este termo indica geralmente duas realidades:

- uma atividade específica (ex. “prova de compreensão oral”);
- um conjunto de atividades (ex. Prova TOEFL ou prova CELPE-Bras).

Neste glossário, a terminologia “prova” se refere à segunda realidade.¹⁸

Sillabo Programma di studio, con una definizione delle competenze attese.¹⁹

Componentes curriculares

Programa de estudo com uma definição das competências esperadas.²⁰

¹⁷ Alla prima accezione qui corrisponde il termine **attività** o “[test] format” o “esercitazione”.

¹⁸ À primeira realidade corresponde o termo **atividade** ou “[teste] format” ou “exercício”.

¹⁹ In italiano si distingue il *sillabo* dal *curricolo* (BALBONI, 1994). Il *sillabo* è il programma di studio, mentre il *curricolo* è un documento più ampio, che contempla anche le scelte metodologiche, gli strumenti e i sussidi, le mete educative e i bisogni degli studenti

²⁰ Em italiano, o *sillabo* (*componentes curriculares*) difere do currículo (BALBONI, 1994). Os *componentes curriculares* correspondem ao programa de estudo, ao passo que o currículo é um

<i>Specifiche</i>	Documento che sintetizza le caratteristiche della prova; vale come guida per l'elaborazione di versioni parallele .
<i>Especificações</i>	Documento que resume as características da prova; é válido como guia para a elaboração de edições posteriores .
<i>Standard</i>	Livello di conoscenza/competenza atteso.
<i>Padrão</i>	Nível de conhecimento/competência esperado.

Strategia di apprendimento linguistico

Piano di azione messo in atto dall'apprendente per sopperire lacune linguistiche e/o ottimizzare le conoscenze di cui dispone.

A dipendere dall'abilità/componente sollecitata, si parla, più specificamente, di *strategie di lettura*, *strategie di scrittura*, *strategie lessicali*, *strategie comunicative*, ecc. (cfr. MACARO, 2010; MACARO *et al.*, 2016).

Estratégia de aprendizagem linguística

Plano de ação posto em prática pelo estudante para suprir lacunas linguísticas e/ou otimizar seus próprios conhecimentos.

Dependendo da habilidade/componente solicitado, falamos, mais especificamente, de *estratégias de leitura*, *estratégias de escrita*, *estratégias lexicais*, *estratégias comunicativas*, etc. (cfr. MACARO, 2010; MACARO *et al.*, 2016).

documento mais amplo, que abrange também as escolhas metodológicas, os instrumentos e os subsídios, os objetivos educacionais e as necessidades dos estudantes.

Strategia metacognitiva Azione intrapresa dall'apprendente al fine di gestire al meglio una serie di **strategie di apprendimento** durante lo svolgimento di una certa attività.

Più in dettaglio la metacognizione riguarda la *pianificazione*, *il monitoraggio* e la *valutazione* degli apprendimenti. Maggiore è la competenza metacognitiva, maggiore è l'autonomia dell'apprendente (CANTOIA *et al.*, 2004; HAUKÅS *et al.*, 2018).

Estratégia metacognitiva

Ação realizada pelo aluno durante uma determinada atividade, a fim de melhor gerir uma série de **estratégias de aprendizagem**.

Em pormenores, a metacognição diz respeito ao *planejamento*, ao *monitoramento* e à *avaliação* da aprendizagem. Quanto maior for a competência metacognitiva, maior será a autonomia do aluno (CANTOIA *et al.*, 2004; HAUKÅS *et al.*, 2018).

Strategia regolativa Azione intrapresa dall'apprendente per gestire i propri stati emotivi (*coltivare l'ottimismo; concedersi una ricompensa a ogni minimo traguardo raggiunto; ecc.*; cfr. ELIAS *et al.*, 2001; MURPHEY, 2006).²¹

Estratégia reguladora Ação realizada pelo aluno a fim de gerir seus próprios estados emocionais (*cultivar o otimismo; dar a si mesmo uma*

²¹ Taluni includono le *strategie regolative* (talora chiamate *affettive*) all'interno della sfera delle **strategie metacognitive** (OXFORD, 1990). In merito all'importanza di questo genere di strategie, scrive la stessa Oxford (1990: 140): "La dimensione affettiva è con ogni probabilità una tra quelle che più incidono sul successo o sul fallimento del processo di apprendimento di una lingua".

*recompensa a cada meta mínima alcançada; etc.; cfr. ELIAS et al., 2001; MURPHEY, 2006).*²²

Strategia di risoluzione di un test

Stratagemma messo in atto dall'allievo per superare una prova, formulato in base alla propria esperienza o acquisito da altri (cfr. A. COHEN, 2006).

Estratégia de resolução de um teste

Tática adotada pelo estudante para passar em uma prova, sendo formulada de acordo com a sua própria experiência ou adquirida de outros (cfr. A. COHEN, 2006).

Versione

(nel contesto di *prove di competenza*) Prova destinata a una certa sessione, la struttura della quale è equipollente rispetto a quella di prove somministrate in altre sessioni, in ottemperanza alle indicazioni delle **Specifiche**.

Versão

(no contexto das *provas de proficiência*) Determinada edição da prova, cuja estrutura deve ser equivalente àquela das provas aplicadas em outras edições, em conformidade com as indicações das **Especificações**.

²² Algumas pessoas incluem as *estratégias reguladoras* (às vezes chamadas *afetivas*) dentro da esfera das **estratégias metacognitivas** (OXFORD, 1990). Acerca da importância de tais estratégias, a própria Oxford escreve (1990: 140): “A dimensão afetiva é, com toda probabilidade, uma das que mais afetam o sucesso ou fracasso do processo de aprendizagem de uma língua”.

B. ELEMENTI

B. ELEMENTOS

Illustriamo una serie di termini che designano aspetti strutturali di una prova.
Apresentamos uma série de termos que designam aspectos estruturais de uma prova.

<i>Batteria</i> (di <i>item</i>)	Serie omogenea di <i>item</i> , riferiti a uno stesso costrutto .
<i>Banco</i> (de <i>itens</i>)	Série homogênea de itens, que se referem ao mesmo constructo .
<i>Chiave</i>	Opzione corretta.
<i>Gabarito</i>	Opção correta.
<i>Descrittore</i>	Presentazione succinta di un livello di competenza o di conoscenza .
<i>Descriptor</i>	Apresentação suscinta de um nível de competência ou de conhecimento .
<i>Distrattore</i>	Opzione non corretta.
<i>Distrator</i>	Opção incorreta.

Griglia di analisi di un testo

(prove di comprensione) Schema utile per accertare l'**adeguatezza** di un testo di lettura/ ascolto in riferimento a un dato livello (cfr. [appendice 2](#)).

Tabela de análise de um texto

(provas de compreensão) Esquema útil para determinar a **pertinência** de um texto de leitura/ escuta, em relação a um determinado nível (cfr. [apêndice 2](#)).

Griglia di valutazione della complessità di un testo

(prove di comprensione) Scheda utile a indagare il grado di complessità di un testo (cfr. TARDIEU *et al.*, 2010; TORRESAN, 2020c).

Tabela de avaliação da complexidade de um texto

(provas de compreensão) Esquema útil para investigar o grau de complexidade de um texto (cfr. TARDIEU *et al.*, 2010; TORRESAN, 2020c).

Griglia di valutazione della performance (Score rubric o Rating scale)

Rappresentazione schematica dei livelli di competenza nelle *abilità di produzione/ interazione/ mediazione* (POLLIT, MURRAY, 1996).

La griglia può essere

- *olistica* (i **descrittori**, rivolti a diversi livelli di competenza, si riferiscono, ciascuno, a più **componenti** del **costrutto**; per esempio, per la produzione scritta, ciascun descrittore riguarda

- al tempo stesso grammatica, lessico e coesione e coerenza);
- *analitica* (i **descrittori**, rivolti a diversi livelli di competenza, si riferiscono, ciascuno, a una singola **componente** del **costrutto**; nel caso della produzione scritta, si darà quindi un descrittore per la grammatica, uno per il lessico e uno per la coesione e coerenza; si vedano gli esempi riportati all'appendice 1).²³

Tabela de avaliação da performance (Score rubric ou Rating scale)

Representação esquemática do nível de competência em habilidades de *produção/ interação/ mediação* (POLLIT, MURRAY, 1996).

A tabela pode ser

- *holística* (os **descritores** indicam diferentes níveis de competência e cada um deles se refere a vários **componentes** do **constructo**; por exemplo, na produção escrita, cada descritor se refere ao mesmo tempo à gramática, léxico e coesão e coerência);
- *analítica* (os **descritores** indicam diferentes níveis de competência e cada um deles se refere a um único **componente** do **constructo**; no caso da produção escrita, teremos um descritor para a gramática, outro

²³ Sulla differenza tra una *griglia olistica* e una *griglia analitica*, cfr. SPINELLI, 2014; MAHAL GALTI *et al.*, 2018. Sul processo di costruzione di una *griglia di valutazione della performance*, cfr. FULCHER *et al.*, 2010. Sulla valutazione della qualità di una *griglia*, cfr. WRIGHT, MASTER, 1982.

para o léxico e outro ainda para coesão e coerência; veja os exemplos no [apêndice 1](#).²⁴

<i>Informazione target</i>	<p>Núcleo informativo al quale è riferito il quesito di una <i>attività di comprensione</i>. Può essere stabilito intuitivamente o mediante mappatura.</p> <p>È bene prevedere una certa distribuzione delle <i>informazioni target</i> all'interno del testo di riferimento (in particolare, per l'ascolto, è necessario evitare <i>item</i> relativi a informazioni troppo ravvicinate o troppo lontane tra loro; una pausa troppo lunga può indurre l'ascoltatore a sentirsi smarrito, nel timore di essersi perso qualcosa).</p>
<i>Informação target</i>	<p>Núcleo informativo da questão em uma <i>atividade de compreensão</i>. Pode ser estabelecido intuitivamente ou mediante mapeamento.</p> <p>É bom prever uma certa distribuição das <i>informações target</i> no texto de referência (especialmente nas atividades de escuta, é necessário evitar itens com informações demasiado próximas ou demasiado afastadas entre si; uma pausa demasiado longa pode fazer com que o ouvinte se sintá desorientado, com medo de ter perdido algo).</p>
<i>Input di appoggio</i>	<p>Si tratta di testi, immagini, grafici o disegni che valgono da stimolo a un'attività di <i>produzione/ interazione/ mediazione</i> orale o scritta.</p>

²⁴ Sobre a diferença entre uma *tabela holística* e uma *tabela analítica*, cfr. SPINELLI, 2014; MAHAL GALTI *et al.*, 2018. Sobre o processo de construção de uma *tabela de avaliação da performance*, cfr. FULCHER *et al.*, 2010. Sobre a avaliação da qualidade de uma *tabela*, cfr. WRIGHT, MASTER, 1982.

<i>Input de apoio</i>	Trata-se de textos, imagens, gráficos ou desenhos que servem de estímulo para uma atividade de <i>produção/ interação/ mediação</i> oral ou escrita.
<i>Istruzioni</i>	Indicazioni rivolte al candidato in merito a cosa gli è richiesto fare. Ai livelli bassi di competenza possono essere formulate nella lingua di origine del candidato.
<i>Instruções</i>	Indicações para o candidato acerca do que ele deve fazer. Em níveis mais baixos de competência, podem ser expressas na língua de origem do candidato.
<i>Item</i>	Quesito a cui è attribuito un punteggio.
<i>Item</i>	Questão à qual é atribuída uma pontuação.
<i>Item âncora</i>	Quesito incluso in versioni parallele di prove, in modo da valere da metro di paragone, ottimizzando la calibrazione .
<i>Item âncora</i>	Questão incluída em versões posteriores de provas, que funciona como medida de comparação, otimizando a adequação .
<i>Item placebo</i>	Quesito molto accessibile (sottodimensionato); è posto all'inizio di una batteria al fine di generare un'aspettativa di successo. In genere, del risultato non si tiene conto.
<i>Item placebo</i>	Questão muito fácil (em relação à competência do aluno); é colocada no início do banco de itens para gerar uma expectativa de sucesso. Em geral, o resultado não é levado em consideração.

Opzioni Alternative tra le quali il candidato è tenuto ad indicare la **chiave**.

Opções Alternativas dentre as quais o candidato deve assinalar a opção correta.

Prompt Stimolo alla *produzione/ interazione/ mediazione* orale o scritta.

Prompt Estímulo à *produção/ interação/ mediação* oral ou escrita.

Punto di taglio (cut score)

Punteggio che definisce la sufficienza.

In alcuni test possiamo avere più punti di taglio, vale a dire punteggi che **discriminano** diversi livelli di competenza (scarso/ insufficiente; insufficiente/ sufficiente; sufficiente/ discreto; discreto/ buono; buono/ ottimo).

Escore de diferenciação (cut score)

Pontuação que define a nota para aprovação.

Em alguns testes podemos ter mais escores de diferenciação, ou seja, pontuações que **discriminam** vários níveis de competência (escasso/ insuficiente; insuficiente/ suficiente; suficiente/ discreto; discreto/ bom; bom/ ótimo).

Scala di valutazione Intervallo di punteggio assegnabile.²⁵

Escala de avaliação Pontuação admissível.²⁶

²⁵ La *scala di valutazione* impiegata in contesto educativo è apparentemente lineare; di fatto è applicata in maniera ordinale, cioè secondo intervalli non congrui. In una scala 0-10, per esempio (cfr. BENVENUTO, 2003),

- è difficile dire che 6 sia il doppio di tre;
- la differenza tra un 5 ½ rispetto a un 6 ha un valore diverso rispetto alla differenza tra un 8 ½ e un 9, nonostante lo scarto sia lo stesso (0.5);
- l'uso invalso di segni aritmetici (6+; 7=; 7≡, ecc.) nella scuola italiana indica la natura discrezionale del giudizio.

In altre parole, una scala di valutazione ordinale consente di stabilire una gerarchia tra dati, un ordine appunto (sappiamo che $6 > 5$), ma non definisce esattamente la differenza tra gli stessi.

²⁶ A *escala de avaliação* usada no contexto educacional é aparentemente linear; na verdade, é aplicada de forma progressiva, ou seja, com intervalos incongruentes. Na escala de 0-10, por exemplo (cfr. BENVENUTO, 2003),

- é difícil dizer que a nota 6 é duas vezes a nota 3;
- a diferença entre um 5,5 em comparação com um 6 não equivale à diferença entre um 8.5 e um 9, embora a margem de diferença entre tais notas seja a mesma (0.5);
- o uso difundido de sinais aritméticos na escola italiana (6+; 7=; 7≡, etc.) indica um critério subjetivo da avaliação.

Em outras palavras, uma escala de avaliação progressiva permite estabelecer uma hierarquia entre os dados, uma ordem, precisamente, (sabemos que $6 > 5$), mas não define exatamente uma diferença entre os mesmos.

C. TIPOLOGIE

C. TIPOLOGIAS

Illustriamo diversi tipi di (a) valutazione, di (b) prova e di (c) attività.

Apresentamos diferentes tipos de (a) avaliação, de (b) prova e de (c) atividade.

A) TIPI DI VALUTAZIONE

A) TIPOS DE AVALIAÇÃO

Autovalutazione

La valutazione è affidata all'apprendente, sollecitato a sviluppare un criterio interno in merito alla qualità della propria prestazione. Si tratta di una modalità diffusa nei libri di testo più recenti, realizzata mediante liste di controllo (“*al termine di questa unità ho imparato a salutare, a chiedere l’ora, ecc.*”).

Una particolare forma di autovalutazione è il **portfolio** (cfr. MARIANI, TOMAI, 2004).

Autoavaliação

A avaliação cabe ao aluno, que é levado a desenvolver um critério interior em relação ao nível de seu desempenho. É um método popular nos livros didáticos mais recentes, implementado através de listas de verificação (“*no final desta unidade aprendi a cumprimentar, a perguntar as horas, etc.*”).

Uma forma específica de autoavaliação é o **portfólio** (cfr. MARIANI, TOMAI, 2004).

Valutazione CLIL (Assessment in Content and Language Integrated Learning)

Il giudizio contempla paritempo l'assimilazione dei contenuti della disciplina oggetto di studio (es. l'architettura, la biologia, ecc.) e la padronanza linguistica (cfr. SERRAGIOTTO, 2006).

Avaliação CLIL (Assessment in Content and Language Integrated Learning)

A avaliação abrange igualmente a assimilação dos conteúdos da matéria em estudo (ex. arquitetura, biologia, etc.) e o domínio linguístico (cfr. SERRAGIOTTO, 2006).

Valutazione collettiva

Oggetto di valutazione è l'esecuzione di un compito da parte di un gruppo. Può essere operata dal docente o da altri gruppi. È praticata nel contesto dell'apprendimento cooperativo (cfr. JOHNSON, JOHNSON, 1999; CASAL, 2016; BENNETT, 2016).

Avaliação coletiva

A avaliação se refere ao desempenho de uma tarefa por um grupo. Pode ser feita pelo professor ou por outros grupos. É praticada no contexto da aprendizagem cooperativa (cfr. JOHNSON, JOHNSON, 1999; CASAL, 2016; BENNETT, 2016).

Valutazione condivisa

Il punteggio viene ripartito tra due o più studenti.

Si pensi a un *role-play*: il voto assegnato è unico e tiene conto della *performance* dei partecipanti (cfr. BAKER, RINVOLUCRI, 2005; MAY, 2009).

In un contesto di classe, l'idea, eventualmente limitata a certe abilità (es. produzione orale) o conoscenze (es. grammatica), ha lo scopo di creare una rete di solidarietà/ cooperazione tra gli studenti (cfr. DAVIS, RINVOLUCRI, 2006).²⁷

Avaliação partilhada

A nota é dividida entre dois ou mais alunos.

Pense em um *role-play*: a nota atribuída é única e leva em conta a performance dos participantes (cfr. BAKER, RINVOLUCRI, 2005; MAY, 2009).

Na sala de aula, a ideia se limita a certas habilidades (ex. produção oral) ou conhecimentos (ex. gramática), e visa criar uma rede de solidariedade/ cooperação entre estudantes (cfr. DAVIS, RINVOLUCRI, 2006).²⁸

Valutazione dinamica

Si raffrontano i risultati raggiunti all'interno di un certo percorso di studio con le conoscenze/ competenze in

²⁷ In una nostra esperienza applicammo la proposta in classi di studenti universitari di italiano LS, dicendo loro “*Chi vuole, forma una coppia con un compagno; il voto assegnato coinciderà con la media delle due prestazioni?*”. La pratica ha indotto all’assunzione di una certa responsabilità da parte dei membri: il più esperto tende ad aiutare il meno esperto; quest’ultimo, al fine di non gravare sul voto finale del compagno, si prepara di più.

²⁸ Em uma de nossas experiências, aplicamos a proposta em turmas de estudantes universitários de italiano LE, dizendo-lhes “*Quem quiser, pode formar uma dupla com um colega; a nota atribuída corresponderá à média das duas avaliações?*”. Tal prática estimulou uma certa responsabilidade por parte dos membros: aquele com melhor desempenho procurou ajudar o de menor desempenho; este, para não prejudicar a nota final daquele, se preparou mais para a avaliação.

entrata, considerando l'evolvere dell'acquisizione per effetto dell'impegno dell'allievo, della mediazione del docente e dell'interazione con i compagni (cfr. LEUNG, 2007; LANTOLF, POEHNER 2010).²⁹

Avaliação dinâmica Comparamos os resultados obtidos dentro de uma determinada trajetória de estudo com os conhecimentos/ competências iniciais, considerando a evolução da aquisição como resultado do empenho do estudante, da mediação do professor e da interação com os colegas (cfr. LEUNG, 2007; LANTOLF, POEHNER, 2010).³⁰

Valutazione a distanza

Modalità di valutazione remota, con esercitazioni sincrone (cfr. NOVELLO, 2021).

Avaliação à distância Modo de avaliação remota, com exercícios síncronos (cfr. NOVELLO, 2021).

Valutazione diffusa (o continua)

Anziché limitata a una esercitazione (a fine corso) o eventualmente a due (una a metà e l'altra a fine corso),

²⁹ Raffrontiamo il progresso di due allievi:

- lo studente A consegue un voto 8 in una prima attività e un voto 9 in una seconda;
- lo studente B consegue un voto 5 in una prima attività e un voto 7.5 in una seconda.

In un senso assoluto la prestazione dello studente A è migliore di quella dello studente B. In termini "dinamici" (cioè di incremento relativo) lo studente B ha ottenuto risultati migliori rispetto allo studente A: + 50% *versus* + 12.5%.

³⁰ Vamos comparar o progresso de dois estudantes:

- o estudante A tira uma nota 8 na primeira atividade e uma nota 9 na segunda;
- o estudante B tira uma nota 5 na primeira atividade e uma nota 7.5 na segunda.

Em termos absolutos, o desempenho do estudante A é melhor do que o do estudante B. Em termos "dinâmicos" (ou seja, de aumento relativo), o estudante B obteve resultados superiores aos do estudante A: + 50% *versus* + 12.5%.

la valutazione è scandita attraverso una serie di osservazioni ripetute, allo scopo di accertare l'assimilazione di segmenti specifici del **sillabo**. Si intende, in questo modo, potenziare l'appreso, facilitando l'**autoregolazione** da parte dell'allievo (McDONOUGH, 2001).³¹

Avaliação contínua

Em vez de se limitar a uma prova (de fim de curso) ou possivelmente duas (uma, na metade e outra, no fim do curso), a avaliação é realizada através de uma série de observações repetidas, a fim de verificar a assimilação de segmentos específicos dos **componentes curriculares**. Pretende-se, desta forma, reforçar a aprendizagem, facilitando a **autorregulação** pelo estudante (McDONOUGH, 2001).³²

Valutazione formativa

Sotto questa dizione consideriamo ogni pratica valutativa (**valutazione dell'interlingua, autovalutazione, valutazione tra pari, valutazione dinamica, valutazione diagnostica**, ecc.) orientata a formare un atteggiamento autonomo nell'apprendente. In generale, l'accento è posto su ciò che lo studente è in grado di fare, piuttosto che su quello che non sa fare (REA-DICKINS, 2006: 168). La *valutazione formativa*, in tal senso, si propone di rettificare le **attribuzioni causali esterne** (BLACK, WILLIAM, 1998).

³¹ Occorre comunque evitare l'eccesso di test (**ipervalutazione**), considerato il disagio che ciò può procurare.

³² No entanto, é necessário evitar o **excesso de avaliações**, dado o desconforto que isso pode causar.

Avaliação formativa Sob este termo, consideramos todas as práticas de avaliação (**avaliação da interlíngua, autoavaliação, avaliação entre pares, avaliação dinâmica, avaliação diagnóstica**, etc.) que geram uma atitude autônoma no aluno. Em geral, a ênfase é colocada naquilo que o estudante é capaz de fazer, ao invés de destacar o que ele não sabe fazer (REA-DICKINS, 2006: 168). Nesse sentido, a *avaliação formativa* visa retificar as **atribuições de causalidade externas** (BLACK, WILLIAM, 1998).

Valutazione informatizzata

Viene realizzata su pc; le eventuali prove di produzione orale vengono registrate, senza che vi sia un'interazione con un intervistatore (per esempio, TOEFL iBT).

Avaliação informatizada

É realizada no computador; as provas de produção oral são gravadas, sem qualquer interação com um entrevistador (por exemplo, TOEFL iBT).

Valutazione dell'interlingua

La teoria dell'*interlingua* prevede che lo studente attraversi una serie di fasi di apprendimento, le caratteristiche delle quali in parte riflettono le lingue di partenza e in parte vi prescindono (NJIEYEP NJOUOGWA, 2007). Le competenze/ conoscenze si assestano grazie a un raffinamento progressivo delle ipotesi da parte dell'allievo, elaborate perlopiù a livello inconscio (cfr. PALLOTTI, 2005).

Avaliação da interlíngua

A teoria da *interlíngua* prevê que o estudante passe por uma série de fases de aprendizagem, cujas características refletem, por um lado, a língua de partida e, por outro, a desconsideram (NJIEYEP NJOUOGWA, 2007). As competências/ os conhecimentos se estabelecem graças a um progressivo refinamento das hipóteses, que são elaboradas principalmente em nível inconsciente, por parte do estudante (cfr. PALLOTTI, 2005).

Valutazione tra pari Anziché dal docente, la valutazione viene operata dai compagni. È applicata in genere alla *produzione orale/ scritta*, e si attua mediante l'uso di una **griglia** semplificata (cfr. appendice 3)³³. Affinché i pareri dei compagni siano affidabili, occorre un previo allenamento, sotto la guida del docente (cfr. MATSUNO, 2009). Il giudizio degli apprendenti può eventualmente essere integrato con quello del docente.³⁴

Avaliação entre pares Em vez do professor, a avaliação é realizada pelos estudantes. É aplicada, geralmente, à *produção oral/ escrita*, e é feita através de uma **tabela** simplificada (cfr. apêndice 3).³⁵ Para que as opiniões sejam confiáveis, é

³³ In sede d'aula, la griglia può essere creata insieme agli apprendenti (CASSANY, 1993; FRAILE *et al.*, 2017).

³⁴ L'efficacia di questo tipo di valutazione è sensibile a fattori socioculturali. Nella cultura cinese, per esempio, nella quale è importante mantenere il senso di armonia del gruppo, ogni forma di giudizio mosso dai pari può essere colta come una critica (CARSON, NELSON, 1996). Di conseguenza, in questo contesto il *feedback* del docente si dimostra più efficace del *feedback* del compagno (CONNOR, ASENAVAGE, 1994).

³⁵ Na sala de aula, a tabela pode ser criada junto com os alunos (CASSANY, 1993; FRAILE *et al.*, 2017).

necessário um treinamento prévio dos estudantes, sob a orientação do professor (cfr. MATSUNO, 2009). A avaliação dos alunos pode ser integrada, eventualmente, com a do professor.³⁶

Valutazione personalizzata

Sono previsti strumenti e **standard** diversi per diversi tipi di studenti (es. studenti con bisogni educativi speciali [DALOISO, 2015]; studenti *gifted* [NOVELLO, 2018]; ecc.).³⁷

Avaliação personalizada

Diferentes ferramentas e **padrões** estão disponíveis para diferentes tipos de alunos (ex. estudantes com necessidades educativas especiais [DALOISO, 2015]; estudantes superdotados [NOVELLO, 2018]; etc.).³⁸

³⁶ A eficácia desse tipo avaliação é afetada por fatores socioculturais. Na cultura chinesa, por exemplo, onde é importante manter o senso de harmonia do grupo, qualquer forma de julgamento pelos pares pode ser lida como uma crítica (CARSON, NELSON, 1996). Consequentemente, neste contexto, o *feedback* do professor se mostra mais eficaz do que o *feedback* do estudante (CONNOR, ASENNAVAGE, 1994).

³⁷ In ambito italiano, cfr. artt. 5 e 6 del Decreto Ministeriale 5669/2011, secondo i quali la valutazione degli alunni con disturbi specifici di apprendimento “dev’essere coerente con gli interventi pedagogico-didattici” inclusi nel “Piano didattico personalizzato”. Il “Piano didattico personalizzato” è un programma ritagliato su misura per l’allievo con un disturbo specifico nell’apprendimento.

³⁸ No contexto italiano, cfr. arts. 5 e 6 do Decreto Ministerial 5669/2011, segundo os quais a avaliação dos alunos com dificuldades de aprendizagem específicas “deve ser coerente com as intervenções pedagógico-didáticas” incluídas no “Plano de Ensino Personalizado”. O “Plano de Ensino Personalizado” é um programa feito sob medida para o aluno com dificuldade de aprendizagem específica.

Valutazione riferita a una norma (o normativa)

L'esito della prestazione di uno studente è definito in rapporto all'esito della prestazione dei compagni.³⁹ Può favorire l'instaurarsi di un clima competitivo (cfr. KOHN, 1993).⁴⁰

Avaliação normativa O resultado do desempenho do estudante é definido em relação ao desempenho dos colegas.⁴¹ Pode incentivar a criação de um clima competitivo (cfr. KOHN, 1993).⁴²

Valutazione riferita a dei criteri (o criteriale)

La prestazione di uno studente è definita in base al raggiungimento di alcuni **standard** (prescinde dalla comparazione tra la prestazione del singolo e quella dei compagni).

Avaliação criterial O desempenho de um aluno é definido de acordo com a aquisição de alguns **padrões** (independentemente da comparação entre o desempenho do indivíduo e o dos colegas).

³⁹ Scrive Novello (2014: 42): “Quando verificiamo i progressi degli studenti e rapportiamo la prestazione di un allievo con quella degli altri, assegnando il punteggio sulla base della prestazione complessiva del gruppo, attuiamo una valutazione normativa; in questo modo abbiamo la possibilità di affermare che uno studente A ha imparato più di uno studente B, ma meno di uno studente C.”

⁴⁰ Per questa ragione in alcuni contesti scolastici il voto è trasmesso all'interessato nella più assoluta riservatezza.

⁴¹ Novello (2014: 42) escreveu: “Quando verificamos o progresso dos estudantes e comparamos o desempenho de um aluno com o dos outros colegas, atribuindo uma nota com base no desempenho geral do grupo, implementamos uma avaliação normativa; isso nos dá a oportunidade de dizer que o estudante A aprendeu mais do que o estudante B, mas menos o que o estudante C.”

⁴² Por essa razão, em alguns contextos escolares, a nota é transmitida ao interessado em absoluto sigilo.

Valutazione sommativa

Raccolta di dati al termine di un percorso di studio (ROCCA, 2016).

Avaliação somativa

Coleta de dados no final de um percurso de estudo (ROCCA, 2016).

B) TIPI DI PROVA

B) TIPOS DE PROVAS

Portfolio

È una documentazione delle prestazioni significative dell'apprendente (*prodotti*), unita a una ricognizione dei modi in cui questi ha svolto i compiti e/o risolto i problemi assegnati (*processi*).

Portfólio

É a documentação do desempenho significativo do aluno (*produto*), que leva em consideração a maneira como ele executou as tarefas e/ou resolveu os problemas atribuídos (*processo*).

Prova per abilità

Si compone di attività volte ad accertare il grado di *competenza* dello studente nella *comprensione* e/o *produzione* di testi.⁴³

Prova de habilidades

Consiste em atividades destinadas a verificar o grau de *competência* do estudante na *compreensão* e/ou na *produção* de textos.⁴⁴

⁴³ Le *conoscenze linguistiche* possono essere valutate indirettamente mediante l'analisi dei testi prodotti.

⁴⁴ Os *conhecimentos linguísticos* podem ser avaliados indiretamente através da análise dos textos produzidos.

Prova di accesso (o di ammissione)

Il superamento consente l'accesso a un percorso di studio o all'esercizio di una certa professione. La prova, in alcuni casi, si riduce a una sola attività (es.: tradurre un breve passaggio dalla LS alla L1).

Prova de ingresso

A aprovação permite o acesso a uma trajetória de estudo ou o exercício de uma determinada profissão. A prova, em alguns casos, se resume a uma única atividade (ex.: traduzir um breve trecho de LE para L1).

Prova adattiva

Vale nel contesto di **valutazioni informatizzate**: vengono automaticamente somministrati *item* più facili o difficili in funzione delle risposte fornite dal *test taker*.

Prova adaptativa

Aplica-se no contexto de **avaliações informatizadas**: itens mais fáceis ou mais difíceis são administrados automaticamente em função das respostas do examinando.

Prova centralizzata

È elaborata da una commissione e viene applicata in più sedi. La centralizzazione ha lo scopo di aumentare l'**affidabilità** della prova (può avere inoltre l'obiettivo di esercitare un **impatto** positivo sull'insegnamento, definendo degli **standard** comuni).

Prova externa

É elaborada por uma comissão e aplicada em várias instituições. A centralização visa aumentar a **confiabilidade** da prova (pode também ter o objetivo de exercer um **impacto** positivo no ensino, definindo **padrões** comuns).

<i>Prova di competenza</i> (o <i>certificazione</i> o <i>proficiency test</i>)	Viene elaborata da enti appositi e si rivolge a un'ampia utenza. Ha lo scopo di attestare una conoscenza linguistica e/o competenza comunicativa in rapporto a determinati livelli. ⁴⁵
<i>Prova de proficiência</i>	É desenvolvida por instituições específicas e se destina a um grande público. Seu objetivo é certificar um conhecimento linguístico e/ou uma competência comunicativa em relação a determinados níveis. ⁴⁶
<i>Prova diagnóstica</i>	È volta a identificare specifiche difficoltà del candidato (cfr. ALDERSON <i>et al.</i> , 2015).
<i>Prova diagnóstica</i>	É desenvolvida para identificar dificuldades específicas do candidato (cfr. ALDERSON <i>et al.</i> , 2015).
<i>Prova di fine ciclo</i>	Rappresenta il conseguimento di alcuni standard alla fine di un ciclo di istruzione e presuppone una preparazione sistematica da parte del candidato in un contesto di educazione formale.
<i>Prova de conclusão</i>	Representa a aquisição de alguns padrões no final de um curso de estudos e requer uma preparação sistemática do candidato em um contexto de educação formal.

⁴⁵ Per una ricognizione delle maggiori certificazioni europee, cfr. Novello, 2009. Per approfondimenti sul CELPE-Bras, cfr. Scaramucci, 2012; Schoffen, Mendel, 2017; Vicentini, 2019.

⁴⁶ Para um levantamento das principais provas de proficiência europeias, cfr. Novello, 2009. Para mais informações sobre CELPE-Bras, cfr. Scaramucci, 2012; Schoffen, Mendel, 2017; Vicentini, 2019.

Prova integrata Ha per oggetto la competenza nell'esercizio di **abilità integrate**. In genere consta di attività di sintesi (esempi: il test di competenza in portoghese CELPE-Bras, o il francese *Diplôme de compétence en langue* [DCL], o l'inglese *Trinity*).

Prova integrada Trata da competência no exercício das **habilidades integradas**. Geralmente, consiste em atividades de síntese (exemplos: a prova de proficiência em português CELPE-Bras, ou em francês *Diplôme de compétence en langue* [DCL], ou em inglês *Trinity*).

Prova di piazzamento (placement test)

È realizzata all'inizio di un corso, in modo da distribuire gli studenti nelle classi a seconda dei livelli di competenza.

Prova de nivelamento

É realizada no início de um curso, a fim de distribuir os alunos em turmas de acordo com os níveis de competência.

Prova di profitto (achievement test)

Verifica se e quanto un alunno ha fatto proprie le conoscenze/ competenze impartite in un certo arco di tempo (è la declinazione concreta di una **valutazione sommativa**).

Una prova di questo tipo fornisce indirettamente all'insegnante indicazioni su

- l'efficacia della metodologia adottata (informa, per esempio, se è stato dedicato tempo sufficiente

- per affrontare un certo tema o per proceduralizzare una certa abilità);
- l'adeguatezza del **sillabo** (avvisa se il programma è stato troppo concentrato, se sono necessarie delle integrazioni, e così via).

Prova de conhecimentos (achievement test)

Verifica se e em que medida o aluno adquiriu os conhecimentos/as competências transmitidas ao longo de um determinado período de tempo (é a forma concreta de uma **avaliação somativa**).

Uma prova deste tipo fornece indiretamente ao professor orientação sobre

- a eficácia da metodologia adotada (informa, por exemplo, se houve tempo suficiente para tratar de um determinado tópico ou para desenvolver uma determinada habilidade);
- a pertinência dos **componentes curriculares** (adverte se o programa foi demasiado compactado, se é necessário complementá-lo, etc).

C) TIPI DI ATTIVITÀ

C) TIPOS DE ATIVIDADES

Attività aperta (o non strutturata)

Comporta una produzione estesa da parte del candidato.

Atividade com resposta livre

Requer uma produção ampla por parte do candidato.

<i>Attività chiusa (o strutturata)</i>	Prevede una scelta tra opzioni .
<i>Atividade com gabarito</i>	Prevê uma escolha entre opções .
<i>Attività integrata</i>	Comporta l'uso di abilità integrate (es. traduzione, riassunto, ecc.).
<i>Atividade integrada</i>	Requer o uso de habilidades integradas (ex. tradução, resumo, etc.).
<i>Attività di coppia</i>	Viene svolta cooperativamente tra due. Si pensi, ad esempio, al <i>role-play</i> .
<i>Atividade em dupla</i>	É realizada cooperativamente entre dois estudantes. Pense, por exemplo, no <i>role-play</i> .
<i>Attività elegibile</i>	Il candidato ha la facoltà di scelta tra più esercitazioni, equivalenti tra loro in termini di difficoltà (cfr. GIPPS, 1994).
<i>Atividade com possibilidade de escolha</i>	O candidato pode escolher entre várias atividades, equivalentes entre si em termos de dificuldade (cfr. GIPPS, 1994).
<i>Attività di gruppo</i>	Viene svolta in gruppo. Si pensi, ad esempio, a una <i>discussione</i> .
<i>Atividade em grupo</i>	É realizada em grupo. Pense, por exemplo, em um <i>debate</i> .
<i>Attività individuale</i>	Viene svolta dal singolo.
<i>Atividade individual</i>	É realizada por um estudante sozinho.

Attività di transcodificazione

(nelle prove di comprensione) Il candidato dimostra di aver compreso eseguendo le istruzioni trasmesse dal testo; per esempio, disegna il paesaggio descritto in un *file* audio.

Atividade de transcodificação

(nas provas de compreensão) O candidato demonstra que compreende executando as instruções transmitidas por um texto; por exemplo, desenhar a paisagem descrita em um *arquivo* de áudio.

Attività semistrutturata

Implica una produzione breve da parte dell'allievo (es. *completamento di frasi; risposte a domande aperte, ecc.*).

Atividade semiestruturada

Requer uma curta produção por parte do estudante (ex. *completar sentenças; responder perguntas abertas, etc.*).

D. PROPRIETÀ

D. PROPRIEDADES

Illustriamo caratteristiche relative a (a) prove o attività, da un lato, e (b) *item*, istruzioni e *prompt*, dall'altro.

Apresentamos características relacionadas a (a) provas ou atividades, por um lado, e (b) item, instruções e *prompt*, por outro.

A) PROPRIETÀ DI UNA PROVA/ ATTIVITÀ

A) PROPRIEDADES DE UMA PROVA/ ATIVIDADE

Accettazione (o accettabilità; *face validity*)

Apprezzamento della prova/attività da parte dei fruitori.
I candidati riconoscono la prova come valida.

Aceitação

Apreço pela prova/atividade por parte dos destinatários.
Os candidatos reconhecem a prova como válida.

Adeguatezza

Un'attività è adeguata quanto più

- il *format* è **familiare** ai candidati;
- le **istruzioni** sono chiare;
- il punteggio è trasparente;
- il testo/ il **prompt** su cui si basa è **calibrato**.

Pertinência

Uma atividade é pertinente quanto mais

- o **tipo de prova é familiar** para os candidatos;

- as **instruções** são claras;
- a nota é transparente;
- o texto/ o **prompt** no qual se baseia está **adequado**.

Affidabilità

Un'attività è tanto più affidabile quanto più è riproducibile, senza alterazioni di risultato (la prestazione dell'alunno si suppone non vari al variare di chi somministra la prova, del momento in cui la prova è somministrata, del luogo in cui la prova è somministrata, ecc.).⁴⁷

In un'attività composta di diversi *item*, il grado di **omogeneità** tra gli *item* concorre all'affidabilità interna della stessa attività (cfr. **Alfa di Cronbach**):⁴⁸ gli *item* cioè svolgono una medesima funzione (è possibile, peraltro, calcolare l'apporto di ogni singolo *item* all'affidabilità interna dell'attività).⁴⁹

⁴⁷ Leggiamo in Grabowski, Dakin (2014: 765): “Se, al variare di una o più condizioni (es. il momento [in cui la prova viene eseguita], le griglie a disposizione del valutatore [...], ecc.), il risultato varia, le informazioni elicitate dalla prova sono inattendibili. In termini ideali, il punteggio ottenuto (definito mediante osservazione) è un'approssimazione dell'effettiva conoscenza/ abilità del candidato (vale a dire del punteggio reale). Più il punteggio [osservato] rispecchia l'effettiva abilità/ conoscenza [cioè, il punteggio reale], maggiore è l'affidabilità della prova”.

⁴⁸ Più elementare rispetto al calcolo dell'**Alfa di Cronbach** è l'analisi “Split half”, che accerta il grado di correlazione tra le due parti in cui è divisa una prova (per esempio tra gli *item* pari, da un lato, e gli *item* dispari, dall'altro).

⁴⁹ Alternativamente, l'*affidabilità* può essere accertata somministrando due volte l'attività a un medesimo campione (*test-retest*), con un lasso di tempo non così ravvicinato da permettere allo studente di ricordare le chiavi né così lungo da prevedere un incremento di competenza (c'è da controrribattere, però, che lo stesso svolgimento dell'esercitazione contribuisce a un aumento della competenza; *test-enhanced learning*).

Un'altra possibilità è misurare la sovrapposizione dei risultati di attività parallele (tuttavia, anche in questo caso, è difficile immaginare attività calibrate millimetricamente su uno stesso grado di difficoltà).

Nell'ambito della valutazione della **performance**, l'affidabilità del valutatore può essere indagata in due modi:

- in termini di conformità del giudizio a quello di altri valutatori (*affidabilità esterna del giudizio del valutatore*);
- in termini di stabilità del giudizio (*affidabilità interna del giudizio del valutatore*).

In sostanza, il giudizio non si discosta eccessivamente dalla media del giudizio degli altri (*affidabilità esterna*) e il valutatore dà prova di essere coerente con se stesso (*affidabilità interna*).⁵⁰

Confiabilidade

Uma atividade é tanto mais confiável quanto mais reprodutível for, sem qualquer alteração no resultado (presume-se que o desempenho do aluno não varie quer com a variação de quem aplica a prova, quer com o momento em que a prova é aplicada, quer com o local em que a prova é aplicada, etc.).⁵¹

Em uma atividade composta por vários itens, o grau de **homogeneidade** entre esses itens contribui para a confiabilidade interna da própria atividade (cfr. **Alpha**

⁵⁰ Per accertare l'*affidabilità interna ed esterna del giudizio del valutatore* si può ricorrere a un'**Analisi Rasch**.

⁵¹ Lemos em Grabowski, Dakin (2014: 765): “Se ocorre alteração em uma ou mais condições (ex. o momento [em que a prova é aplicada], as tabelas à disposição do avaliador [...], etc.), e o resultado varia, as informações obtidas pela prova não são confiáveis. Em termos ideais, a nota obtida (definida pela observação) é uma aproximação do conhecimento/ da habilidade real do candidato (ou seja, a nota real). Quanto mais a nota [observada] reflete a habilidade/o conhecimento real [ou seja, a nota real], maior é a confiabilidade da prova”.

de Cronbach):⁵² os itens desempenham uma mesma função (além disso, é possível calcular a contribuição de cada item para a confiabilidade interna da atividade).⁵³ No âmbito da avaliação da **performance**, a confiabilidade do avaliador pode ser investigada de duas formas:

- em termos de consonância com a nota atribuída por outros avaliadores (*confiabilidade externa da avaliação do avaliador*);
- em termos de estabilidade da avaliação (*confiabilidade interna da avaliação do avaliador*).

Basicamente, a avaliação não diverge excessivamente da média da avaliação de outros (*confiabilidade externa*) e o avaliador revela-se coerente consigo mesmo (*confiabilidade interna*).⁵⁴

⁵² Mais elementar que o cálculo de **Alfa de Cronbach** é a análise “Split half”, que verifica o grau de correlação entre as duas partes em que uma prova está dividida (por exemplo entre os itens pares, por um lado, e os itens ímpares, por outro).

⁵³ Alternativamente, a confiabilidade pode ser determinada aplicando duas vezes a atividade para o mesmo grupo (*test-retest*), com um intervalo de tempo nem tão curto, que permitiria ao estudante lembrar o gabarito, nem tão longo, pois permitiria o aumento da competência (deve-se assinalar, no entanto, que o exercício em si contribui para um aumento da competência; *test-enhanced learning*).

Outra possibilidade é comparar os resultados de atividades equivalentes (no entanto, mesmo neste caso, é difícil imaginar atividades reguladas milimetricamente, com o mesmo grau de dificuldade).

⁵⁴ A fim de verificar a *confiabilidade interna e externa da nota atribuída pelo avaliador*, pode ser utilizada a **Análise Rasch**.

Autenticità (del testo o di un'attività)

L'attributo *autentico* può riferirsi

- al testo oggetto di comprensione (un testo autentico si oppone a un testo scritto *ad hoc* o **manipolato**);
- a un'attività, nella misura in cui
 - o la stessa riflette le operazioni cognitive di cui l'allievo dà prova in una situazione di uso naturale della lingua;
 - o il contesto evocato riflette un contesto di uso spontaneo della lingua⁵⁵

Autenticidade (do texto ou de uma atividade)

O atributo *autêntico* pode fazer referência

- ao texto de leitura/escuta (um texto autêntico se opõe a um texto escrito *ad hoc* ou reescrito [cfr. **Reelaboração**]);
- a uma atividade, na medida em que
 - o esta reflete as operações cognitivas que o estudante demonstra em uma situação de uso natural da língua;
 - o o contexto simulado reflete um contexto de uso espontâneo da língua⁵⁶

⁵⁵ Bachman (1990) parla rispettivamente di *interactional authenticity* (una sorta di autenticità cognitiva) e di *situational authenticity* (il contesto evocato simula un contesto naturale). Per approfondimenti sul concetto di *autenticità* nell'ambito della valutazione linguistica, cfr. Shohamy, Reves, 1985.

⁵⁶ Bachman (1990) fala respectivamente de *interactional authenticity* (uma espécie de autenticidade cognitiva) e de *situational authenticity* (o contexto simulado se aproxima de um contexto natural). Para mais informações sobre o conceito de *autenticidade* na avaliação linguística, cfr. Shohamy, Reves, 1985.

Calibrazione (o dimensionamento)

Una prova, un'attività o un testo sono calibrati quando il livello di difficoltà è rapportato al livello di competenza del candidato.

Adequação

Uma prova, uma atividade ou um texto estão adequados quando o nível de dificuldade é proporcional ao nível de competência do candidato.

Completezza (esaustività)

Proprietà di una prova, definita dalla combinazione della **copertura** e di un'adeguata **estensione**.

Exaustividade

Propriedade de uma prova, definida pela combinação de **abrangência** e **tamanho** adequado.

Convergenza

(per una prova di profitto) La prova riflette il *focus* del corso. Se, per esempio, il *focus* è comunicativo, è inappropriato servirsi di una prova esclusivamente o prevalentemente grammaticale.

Convergência

(para uma prova de conhecimentos) A prova reflete o foco do curso. Por exemplo, se o foco é comunicativo, é inadequado usar uma prova exclusivamente ou predominantemente gramatical.

Copertura

Ampia varietà di tipologie di attività all'interno di una prova, al fine di considerare le varie **componenti** relative al **costrutto di riferimento**.⁵⁷

⁵⁷ A differenza dell'**estensione**, che si riferisce a un criterio quantitativo, la *copertura* rappresenta un concetto qualitativo.

<i>Abrangência</i>	Grande variedade de tipos de atividades dentro de uma prova, de modo a considerar os diversos componentes relacionados ao constructo de referência . ⁵⁸
<i>Discriminatività</i>	Una prova/ attività discrimina quando è in grado di distinguere gli studenti più abili dai meno abili (vedi relativi indici).
<i>Discriminação</i>	Uma prova/ atividade discrimina quando é capaz de distinguir os estudantes mais capazes daqueles menos capazes (ver índices relacionados).
<i>Durata</i>	Tempo concesso all'apprendente per svolgere l'attività/ la prova. Il calcolo dei tempi, da parte del progettista, si rivela di estrema importanza: margini troppo stretti generano ansia, la quale influisce negativamente sulla prestazione.
<i>Duração</i>	Tempo concedido ao aluno para realizar a atividade/ a prova. O cálculo do tempo, por parte de quem projeta a prova, é extremamente importante: pouco tempo disponível gera estresse, o que afeta negativamente o desempenho.
<i>Economia</i>	Rapporto positivo tra costi e benefici in riferimento all'elaborazione/ somministrazione di una prova/ un'attività.

⁵⁸ Ao contrário do **tamanho**, que remete a um critério quantitativo, a *abrangeência* é um conceito qualitativo.

Economia

Custo-benefício positivo em relação à elaboração/ aplicação de uma prova/ uma atividade.

Elicitazione (test method effect)

Ogni genere di attività sollecita un comportamento specifico (cfr. TSAGARI, 1994). Per esempio, nell'ambito delle attività di comprensione, un *quesito a scelta multipla* si accompagna a una lettura/ un ascolto analitico; diversamente, una *domanda aperta* è uno strumento più flessibile: può riguardare tanto le idee principali quanto i dettagli.

Al fine di considerare l'ampia gamma di comportamenti strategici facenti capo a un certo **costrutto** di riferimento, chi allestisce una prova è tenuto a prevedere una certa varietà di attività (**copertura**).

Comportamento específico gerado por uma atividade

Qualquer forma de atividade acarreta um comportamento específico (cfr. TSAGARI, 1994). Por exemplo, nas atividades de compreensão, uma *questão de múltipla escolha* leva a uma leitura/ uma escuta analítica; diferentemente, uma *pergunta aberta* é uma ferramenta mais flexível: pode abranger tanto as ideias principais quanto os detalhes.

A fim de considerar a ampla gama de comportamentos estratégicos que faz parte de um determinado **constructo** de referência, aqueles que elaboram uma prova devem prever uma certa variedade de atividades (**abrangência**).

Estensione

Ampiezza della prestazione richiesta, definita a seconda dell'attività (es. il numero di *item* di cui si compone un esercizio grammaticale; il numero di parole previste in una composizione scritta; la lunghezza di un brano di lettura; la durata di un testo di ascolto; ecc.). Occorre che il comportamento sollecitato sia sufficientemente rappresentativo (ovvero, la prestazione sia adeguatamente estesa), affinché si possa formulare un giudizio attendibile (**estrapolazione**) sulla conoscenza/ competenza oggetto di valutazione (prove/ attività che implicino prestazioni troppo estese vanno, d'altra parte, ridimensionate, posto che generano **affaticamento**).

Tamanho

Desempenho exigido, definido de acordo com a atividade (ex. o número de itens que consta em um exercício gramatical; o número de palavras previstas em uma produção escrita; a extensão de um trecho para leitura; a duração de um texto de escuta; etc.). É necessário que o comportamento solicitado seja sufficientemente representativo (ou seja, que o desempenho seja adequadamente longo) e permita formular uma verificação confiável (**extrapolação**) do conhecimento/ competência a ser avaliado (provas/ atividades que demandam um desempenho muito longo devem, em todo caso, ser redimensionadas, à medida que causam fadiga).

<i>Familiarità</i>	Dimestichezza, da parte dell'apprendente, con le caratteristiche e gli elementi della prova/ dell'attività. ⁵⁹
<i>Familiaridade</i>	Familiaridade do aluno com as características e elementos da prova/ da atividade. ⁶⁰
<i>Impatto (washback)</i>	Conseguenze che la prova esercita sulla didattica (WALL, ALDERSON, 1993; SCARAMUCCI, 2004; MACCIÒ, 2017).
<i>Impacto</i>	Efeito que a prova exerce na didática (WALL, ALDERSON, 1993; SCARAMUCCI, 2004; MACCIÒ, 2017).
<i>Praticabilità (o logistica)</i>	Risorse a disposizione del somministratore, in accordo con il tipo di prova.
<i>Logística</i>	Recursos disponíveis para quem aplica a prova, de acordo com o tipo de prova.
<i>Sicurezza</i>	Tutela della segretezza della prova, in maniera tale che non vengano trapelate informazioni sensibili prima della somministrazione.
<i>Segurança</i>	Proteção do sigilo de prova, de modo que não haja vazamento de informações confidenciais antes da aplicação.

⁵⁹ La non familiarità tende a incidere negativamente sulla prestazione, agendo da variabile esterna al **costrutto di riferimento**.

⁶⁰ A falta de familiaridade tende a afetar negativamente o desempenho, atuando como variável externa ao **constructo de referência**.

Specificità (di un'attività)

(per le attività tese a valutare la conoscenza di componenti linguistiche) L'attività ha per oggetto un tipo di conoscenza linguistica preciso (es. grammaticale), senza valutare al tempo stesso altri tipi di conoscenza linguistica (es. lessicale); ciò concorre alla **validità** e all'**affidabilità** del giudizio.

Especificidade (de uma atividade)

(para atividades destinadas a avaliar o conhecimento de componentes linguísticos) A atividade tem por objeto um tipo de conhecimento linguístico específico (ex. gramatical), sem avaliar ao mesmo tempo outros tipos de conhecimentos linguísticos (ex. lexical); isso contribui para a **validade** e para a **confiabilidade** da avaliação.

Trasparenza

Una prova può avere diversi gradi di trasparenza. Prima che sia somministrata, i candidati possono essere informati in merito a (NOVELLO, 2012)

- i parametri di valutazione;
- il tipo di punteggio;
- le risorse a cui possono accedere durante la prova (es. il dizionario);
- prove già somministrate, sulle quali si possono esercitare.

Successivamente allo svolgimento della prova, i candidati possono ricevere un **feedback** dettagliato in merito alla qualità della prestazione eseguita.

Transparência

Uma prova pode ter diferentes graus de transparência. Antes de realizá-la, os candidatos podem ser informados sobre (NOVELLO, 2012)

- os parâmetros da avaliação;
- o tipo de nota;
- os recursos aos quais podem ter acesso durante a prova (ex. dicionário);
- provas já aplicadas, com as quais podem treinar.

Após a realização da prova, os candidatos podem receber um **feedback** detalhado sobre o seu desempenho.

Validità

La validità di una prova può essere definita sotto diversi aspetti.⁶¹ Diamo menzione ai due di maggior rilievo:

- una prova è valida posto che misuri effettivamente ciò che dichiara di misurare (*validità di costrutto*). Per dire, una prova di scrittura che si componga di attività di reimpiego morfosintattico non può dirsi valida;⁶²
- una prova è valida posto che il contenuto a cui si riferisce è rappresentativo del **dominio** in cui il candidato opera, in un contesto di uso spontaneo della lingua (*validità di contenuto*). Una prova volta ad accertare l'abilità di lettura di testi accademici,

⁶¹ Per una ricognizione del concetto di validità, si considerino Chapelle, 1988; Messick, 1989; Weir, 2005; Jamieson, 2011; D'Este, 2012; Borsboom, Markus, 2013, Fulcher, 2015.

⁶² Il concetto si sovrappone a quello di *interaccional authenticity*, suggerito da Bachman, 1990.

per esempio, deve basarsi su testi effettivamente somministrati in ambito universitario.⁶³

Va precisato, in ogni caso, che dagli anni '80 in poi, anziché essere attribuita alla prova in sé, la validità tende ad essere attribuita all'uso che si fa della prova e all'interpretazione dei risultati (cfr. COOMBE, 2018: 43).⁶⁴ In altre parole, l'aggettivo “valido” qualifica il giudizio.

Validade

A validade de uma prova pode ser definida em vários aspectos.⁶⁵ Mencionamos os dois mais significativos:

- uma prova é válida se medir efetivamente o que afirma medir (*validade de constructo*). Uma prova de escrita, por exemplo, que proponha atividades de análise morfossintática não pode ser considerada válida;⁶⁶
- uma prova é válida desde que o conteúdo a que se refere seja representativo do **domínio** em que o candidato atua, em um contexto de uso espontâneo da língua (*validade de conteúdo*). Uma prova para verificar a habilidade de leitura de textos acadêmicos, por exemplo, deve ser baseada em textos efetivamente utilizados no âmbito universitário.⁶⁷

⁶³ Il concetto si sovrappone a quello di *situational authenticity*, suggerito da Bachman, 1990.

⁶⁴ Messick (1989: 13) scrive “la validità si riferisce a un giudizio in grado di accertare – sulla base di dati empirici e di riflessioni che provengono dalla ricerca – se le interpretazioni formulate e le azioni intraprese a seguito di un test (o di un'altra forma di valutazione) sono appropriate ed efficaci”.

⁶⁵ Para aprofundar o conceito de validade, consultar Chapelle, 1988; Messick, 1989; Weir, 2005; Jamieson, 2011; D'Este, 2012; Borsboom, Markus, 2013, Fulcher, 2015.

⁶⁶ O conceito se sobrepõe ao de *interaccional authenticity*, sugerido por Bachman, 1990.

⁶⁷ O conceito se sobrepõe ao de *situational authenticity*, sugerido por Bachman, 1990.

Note-se, em todo caso, que desde a década de 1980, em vez de ser atribuída à prova em si, a validade tende a ser atribuída ao uso da prova e à interpretação dos resultados (cfr. COOMBE, 2018: 43).⁶⁸ Em outras palavras, o adjetivo “válido” qualifica a atribuição da nota.

B) PROPRIETÀ DI PROMPT, ISTRUZIONI, ITEM

B) PROPRIEDADES DE PROMPT, INSTRUÇÕES, ITENS

Accessibilità (di istruzioni, *item*, *prompt*)

Adeguamento degli elementi che compongono la prova alle esigenze di candidati con bisogni educativi speciali.

Accessibilidade

Adaptação dos elementos da prova às demandas de candidatos com necessidades educativas especiais.

Calibrazione (di un *item*)

Un *item* è calibrato (o dimensionato) quando il livello di difficoltà è rapportato al livello di competenza dello studente.

Adequação (de um *item*)

Um item está adequado quando o nível de dificuldade é proporcional ao nível de competência do estudante.

Discriminatività (di un *item*)

Un *item* discrimina quando è in grado di distinguere gli studenti più abili dai meno abili (vedi relativi **indici**).

⁶⁸ Messick (1989: 13) escreve “a validade refere-se à atribuição de nota capaz de determinar – com base em dados empíricos e reflexões decorrentes da pesquisa – se as interpretações feitas e as medidas tomadas depois de um teste (ou de outra forma de avaliação) são adequadas e eficazes”.

Discriminação (de um item)

Um item discrimina quando é capaz de distinguir os mais capazes daqueles menos capazes (ver **índices** relacionados).

Concisione (di istruzioni e item)

Essenzialità ed immediatezza. Istruzioni poco chiare generano **ansia** nel candidato (MADSEN, 1982).

Concisão (de instruções e itens)

Essencialidade e simplicidade. Instruções pouco claras geram **estresse** no candidato (MADSEN, 1982).

Generatività (di un prompt)

Un *prompt* è generativo nella misura in cui stimola il candidato ad esprimersi (cfr. appendice 4).

Produttivade (de um prompt)

Um *prompt* é produtivo na medida em que estimula o candidato a se expressar (cfr. apêndice 4).

Indipendenza (di un item)

La risoluzione dell'*item* non dipende dalla risoluzione di altri *item* (il candidato cioè non ha modo di indovinare quale sia la risposta sulla base delle informazioni attinte da altri *item*).
L'indipendenza concorre all'**affidabilità** dell'attività.

Independência (de um item)

A resolução do item não depende da resolução de outros itens (o candidato não tem como adivinhar a resposta com base nas informações extraídas de outros itens).
A independência contribui para a **confiabilidade** da atividade.

Omogeneità Un *item* opera in sintonia con gli *item* facenti parte di una stessa **batteria**, facendo riferimento a un medesimo **costrutto**.

L'omogeneità concorre all'**affidabilità** dell'attività.

Homogeneidade Um item funciona em sintonia com itens que fazem parte do mesmo **banco de itens**, referindo-se ao mesmo **constructo**.

A homogeneidade contribui para a **confiabilidade** da atividade.

Specificità (di un item)

(per *item* relativi ad attività volte a valutare la conoscenza di componenti linguistiche) L'*item* ha per oggetto solo un certo tipo di conoscenza linguistica (es. grammaticale), senza valutare al tempo stesso altri tipi di conoscenza linguistica (es. lessicale).

Especificidade (de um item)

(para itens em atividades destinadas a avaliar o conhecimento de componentes linguísticos) O item tem por objeto apenas um certo tipo de conhecimento linguístico (ex. gramatical), sem ao mesmo tempo avaliar outros tipos de conhecimentos linguísticos (ex. lexical).

E. OPERAZIONI

E. PROCEDIMENTOS OPERACIONAIS

Illustriamo una serie di processi che accompagnano un intervento valutativo.
Apresentamos uma série de procedimentos que acompanham uma avaliação.

Allestimento degli spazi

Controllo dell'idoneità dei locali e del distanziamento tra le postazioni.⁶⁹

Organização do espaço

Verificação da infraestrutura das salas e do espaçamento entre o mobiliário.⁷⁰

Allineamento

Armonizzazione dei pareri degli esaminatori in sede di valutazione di una produzione scritta/ orale. La divergenza tra i giudizi viene resa nota e analizzata; nell'arco di più sessioni, gli scarti possono ridursi e le valutazioni diventano più omogenee (cfr. SHOHAMY *et al.*, 1992; CUSHING WEIGLE, 1994; LUMLEY, 2002; LIM, 2011).

⁶⁹ Nel caso di esercitazioni di comprensione orale, la mancanza di una adeguata insonorizzazione degli spazi o l'assenza di casse che consentano una adeguata diffusione del suono possono agire negativamente sulla prestazione dei candidati.

⁷⁰ No caso de exercícios de compreensão oral, a falta de isolamento acústico adequado nos locais e a ausência de alto-falantes que permitam uma difusão adequada do som podem afetar negativamente o desempenho dos candidatos.

Convergência nas avaliações

Aproximação do parecer dos examinadores ao avaliar uma produção escrita/ oral. A divergência nas avaliações é debatida e analisada; ao longo de várias reuniões, as diferenças podem diminuir e as avaliações se tornam mais homogêneas (cfr. SHOHAMY *et al.*, 1992; CUSHING WEIGLE, 1994; LUMLEY, 2002; LIM, 2011).

Assegnazione del punteggio

Conferimento del punteggio sulla base di una determinata **scala**.⁷¹

Atribuição da pontuação

Atribuição da nota de acordo com uma determinada **escala**.⁷²

Comunicazione dei risultati

Trasmissione dei giudizi ai candidati.

Divulgação dos resultados

Divulgação das notas para os candidatos.

⁷¹ In attività grammaticali che prevedano la formulazione di costruzioni complesse si consiglia un doppio punteggio (PURPURA, 2004). Ad esempio, per un tempo composto (come il passato del passato: “avevo pensato”, in italiano, “tinha pensado” in portoghese, “I had thought” in inglese), si assegna un punto per la corretta coniugazione dell’ausiliare e un punto per la resa corretta del verbo principale (nell’esempio, il verbo principale è volto al participio passato).

⁷² Nas atividades gramaticais que envolvem a formulação de construções complexas é recomendada uma dupla pontuação (PURPURA, 2004). Por exemplo, para um tempo composto (como o passado do passado: “avevo pensato”, em italiano, “tinha pensado”, em português, “I had thought”, em inglês), um ponto é atribuído para a conjugação correta do verbo auxiliar e outro ponto para a forma nominal correta do verbo principal (nos exemplos, os verbos estão no participio passado).

<i>Formattazione</i>	Impaginazione dell'attività/ prova.
<i>Diagramação da prova</i>	Layout da atividade/ prova.
<i>Generalizzazione</i>	Estensione a un'intera popolazione di <i>pattern</i> rinvenibili nei risultati di un test applicato a un campione rappresentativo, una volta accertata la significatività statistica .
<i>Generalização</i>	Os padrões encontrados nos resultados de um teste aplicado por amostragem são declarados válidos para toda uma população, levando em consideração a significância estatística .
<i>Manipolazione del testo</i>	(per le esercitazioni di comprensione) Manomissione/ adattamento di fonti.
<i>Reelaboração do texto</i>	(para provas de compreensão) Alteração de um texto para fins de verificação.
<i>Mappatura del testo</i>	(per le attività di comprensione) Rinvenimento delle informazioni target mediante consultazione tra esperti (SARIG, 1989; GREEN, 2017) o di un gruppo rappresentativo di studenti (TORRESAN, 2020b).
<i>Mapeamento do texto</i>	(para atividades de compreensão) Obtenção das informações target mediante parecer de especialistas

(SARIG, 1989; GREEN, 2017) ou de um grupo representativo de estudantes (TORRESAN, 2020b).

Ponderazione (o assegnazione di un peso)

Attribuzione di un punteggio maggiore a un determinato **item** rispetto ad altri *item* (poiché più difficile), o a una determinata **componente** rispetto ad altre componenti (posto che essa abbia una certa salienza nel **costrutto operativo**), o a una determinata attività rispetto ad altre attività (considerata più rappresentativa di altre in riferimento al **dominio** a cui la prova si riferisce).

Peso

Atribuição de uma pontuação mais elevada a um dado **item** do que a outros itens (porque é mais difícil), ou a um dado **componente** do que a outros componentes (desde que tenha uma certa relevância no **constructo operacional**), ou a uma atividade específica sobre outras atividades (considerada mais representativa do que outras, em relação ao **domínio** a que a prova faz referência).

Pre-testing

(in un contesto certificatorio) Analisi dei risultati di una o più attività somministrate a un gruppo pilota (min 30). È importante che il gruppo sia rappresentativo (rifletta cioè le caratteristiche della popolazione a cui è/sono rivolta/e la/le attività). L'analisi si compendia di una fase di **revisione** degli *item*.

Pré-testagem

(em um contexto de provas de proficiência) Análise dos resultados de uma ou mais atividades aplicadas em grupo

de teste (30 pessoas, no mínimo). É importante que o grupo seja representativo (ou seja, deve refletir as características do público para o qual as atividades estão direcionadas). A análise passa por uma fase de **revisão** dos itens.

Post-testing (in un contesto certificatorio) Analisi dei risultati a seguito della **somministrazione** della prova.

Pós-testagem (em um contexto de provas de proficiência) Análise dos resultados após a **aplicação** da prova.

Recupero (nel contesto delle prove di profitto) Opportunità concessa di una verifica supplementare.

Prova de recuperação (no contexto das provas de conhecimento) Possibilidade de realizar uma segunda chamada.

Revisione (o riscrittura)
Raffinamento di *item/ istruzioni/ prompt* a seguito del **pre-testing**.

Controle e revisão das questões da prova
Aperfeiçoamento de *itens/ instruções/ prompt* após a **pré-testagem**.

Scelta del testo (attività di comprensione) Il brano deve andare incontro agli interessi dei candidati ed essere **calibrato**. Le proprietà testuali (*in primis* efficacia, coerenza e coesione) devono essere rispettate. Nel caso specifico dell'audio, il brano dev'essere di buona qualità.

<i>Escolha do texto</i>	(atividades de compreensão) O trecho de leitura deve atender aos interesses dos candidatos e ser adequado . As propriedades textuais (principalmente eficácia, coerência e coesão) devem ser respeitadas. No caso específico do áudio, o texto de escuta deve ter uma boa qualidade técnica.
<i>Scelta del format</i>	Il tipo di attività prescelto deve andare incontro al profilo dei candidati (bisogni, interessi, ecc.). In un contesto certificatorio, al fine di valutare una certa competenza/ abilità, si ricorre a diversi tipi di attività (copertura).
<i>Escolha do tipo de prova</i>	O tipo de atividade escolhida deve atender ao perfil dos candidatos (necessidades, interesses, etc.). Em um contexto de provas de proficiência, utilizam-se diferentes tipos de atividades (abrangência) para avaliar uma determinada competência/ habilidade.
<i>Scrittura degli item</i>	Confezione degli <i>item</i> di cui si compone una certa attività.
<i>Redação dos itens</i>	Preparação dos itens que compõem uma determinada atividade.
<i>Scrittura del testo</i>	(attività di comprensione orale) Ai livelli A1-A2, considerata la difficoltà di reperire materiale autentico con sufficiente contestualizzazione e di qualità (cfr. BRINDLEY, 1998) e, al tempo stesso, nell'ottica di

garantire una equivalenza tra i testi (a vantaggio dell'**affidabilità**), è diffusa la pratica di scrivere degli *scripts*, soggetti poi a interpretazione.⁷³

Redação do texto

(atividades de compreensão oral) Nos níveis A1-A2, dada a dificuldade de encontrar **material autêntico** com contextualização e qualidade suficientes (cfr. BRINDLEY, 1998) e, ao mesmo tempo, com a intenção de assegurar a equivalência entre textos (em benefício da **confiabilidade**), é comum a prática da escrita de *scripts*, os quais são submetidos posteriormente à interpretação.⁷⁴

Simulazione

Modello di prova su cui il candidato può esercitarsi autonomamente o che viene applicato in aula in sede di preparazione.

Simulado

Modelo de prova que funciona como treinamento (o candidato pode treinar de maneira autônoma ou o modelo pode ser aplicado em sala de aula).

Somministrazione

Applicazione della prova o di una esercitazione specifica.

Aplicação

Aplicação de prova ou de exercício específico.

⁷³ Vale comunque la pena di sottolineare il rischio di giungere a versioni troppo 'pulite' da un punto di vista conversazionale, con una sottorappresentazione del **costrutto** (cfr. OACKEY, WAGNER, 2018; PAULETTO, 2020).

⁷⁴ Vale ressaltar, no entanto, o risco de se chegar a versões demasiado 'limpas' do ponto de vista conversacional, com uma sub-representação do **constructo** (cfr. OACKEY, WAGNER, 2018; PAULETTO, 2020).

Standard Setting Ciclo di sessioni in cui gli esperti si confrontano in merito al livello di difficoltà delle attività alla luce dei **descrittori** relativi alla competenza (cfr. CIZEK, BUNCH, 2006; FIGUERAS, NOIJONS, 2009).⁷⁵

Standard Setting Reuniões em que os especialistas debatem o nível de dificuldade das atividades, tendo em vista os **descritores** da competência (cfr. CIZEK, BUNCH, 2006; FIGUERAS, NOIJONS, 2009)⁷⁶

Sviluppo di un testo orale a partire da una traccia

(durante la confezione di un'attività di ascolto) Dialoghi o narrazioni improvvisati sulla base di un canovaccio essenziale.

Preparação com base em esquema do texto oral de prova

(durante a preparação de atividade de escuta) Diálogos e narrativas improvisadas com base em um esquema simples.

⁷⁵ Analogo allo *Standard Setting* è il *benchmarking*: anziché alle attività, esso è relativo alle *performance* (orali/scritte) degli allievi.

⁷⁶ O *benchmarking* é análogo ao *Standard Setting*: em vez das atividades, está relacionado com a *performance* (oral/escrita) dos alunos.

F. ISTITUZIONI E INDIVIDUI

F. PESSOAS E INSTITUIÇÕES

Ci riferiamo qui alle figure/ istituzioni coinvolte nel processo valutativo.

Citamos, a seguir, as pessoas/ instituições envolvidas no processo de avaliação.

Centro d'esame Sede convenzionata in cui viene applicata una **prova di competenza**.

Ente credenciado Local onde a **prova de proficiência** é aplicada.

Centro Certificatore Ente che provvede all'elaborazione di **prove di competenza**.

Ente certificador Ente responsável pela elaboração das **provas de proficiência**.

Correttore Colui/ei che valuta prove di produzione scritta.
Al fine di giungere a un giudizio quanto più attendibile, una stessa composizione può essere valutata indipendentemente da due correttori, di cui sia già accertato un certo grado di **affidabilità** (cfr. BRITTON *et al.*, 1966; WOOD, QUINN, 1976). In caso di disaccordo significativo (con uno scarto superiore al 20%), può essere chiamato in causa un terzo

correttore, e si considera la media tra i due punteggi più prossimi (SHAW, WEIR, 2007).

Corretor de redação

É quem avalia as provas de produção escrita.

Para chegar a uma avaliação mais confiável, a mesma redação pode ser avaliada, de forma independente, por dois corretores, cujo grau de **confiabilidade** já esteja estabelecido (cfr. BRITTON *et al.*, 1966; WOOD, QUINN, 1976). Em caso de discordância significativa (com uma diferença de mais de 20%), um terceiro corretor pode ser convocado, e se respeita a média entre as duas notas mais próximas (SHAW, WEIR, 2007).

Esaminatore

Colui/ei che valuta la *performance*, servendosi in genere di una **griglia di valutazione**.⁷⁷

In sede certificatoria, nell'ambito della valutazione dell'orale, la figura dell'esaminatore è distinta da quella dell'**intervistatore**.

Examinador

É quem avalia a *performance*, normalmente usando uma **tabela de avaliação**.⁷⁸

No âmbito da proficiência, e no tocante à avaliação oral, o papel do examinador é distinto do papel do **entrevistador**.

⁷⁷ Nella valutazione dello scritto, il termine può essere sostituito da quello di **correttore**.

⁷⁸ Na avaliação da escrita, o termo pode ser substituído pelo termo **corretor de redação**.

<i>Falso positivo</i>	Candidato che ottiene un risultato positivo, superiore a quello che in realtà sa/sa fare.
<i>Falso positivo</i>	Candidato que obtém um resultado positivo, maior do que aquilo que ele realmente sabe/sabe fazer.
<i>Falso negativo</i>	Candidato che ottiene un risultato negativo, inferiore a quello che in realtà sa/sa fare.
<i>Falso negativo</i>	Candidato que obtém um resultado negativo, inferior ao que ele realmente sabe/sabe fazer.
<i>Giudice</i>	Esperto che partecipa a sessioni di Standard Setting .
<i>Juíz</i>	Especialista que participa das reuniões de Standard Setting .
<i>Inlier</i>	Candidato che manifesta comportamenti inattesi nella risoluzione di <i>item</i> o nello svolgimento di <i>task</i> calibrati al proprio livello di competenza (per distrazione, per esempio).
<i>Inlier</i>	Candidato que apresenta comportamentos inesperados na resolução de itens ou na execução de tarefas adequadas ao seu nível de competência (por distração, por exemplo).
<i>Intervistatore</i>	Esperto che interagisce con il candidato durante attività di produzione orale.
<i>Entrevistador</i>	Especialista que interage com o candidato durante as atividades de produção oral.

Item writer (o estensore/ autore dell'esercitazione)

Persona incaricata della redazione degli *item*.⁷⁹

Redator de itens

É o responsável pela redação dos itens.⁸⁰

Outlier

Candidato che manifesta comportamenti inattesi rispetto al campione; i risultati si collocano alle estremità della curva di distribuzione.

Outlier

Candidato que apresenta comportamentos inesperados em relação ao grupo; os resultados se situam nas extremidades da curva de distribuição.

Moderatore

Colui/ei che conduce sessioni di **Standard Setting**.

Moderador

É quem norteia as reuniões de **Standard Setting**.

*Test taker (o candidato)*⁸¹

Colui/ei che si sottopone a una prova/ esercitazione.

*Examinando (ou candidato)*⁸²

É quem é testado pela prova/ atividade.

⁷⁹ Spesso il termine viene usato in maniera più ampia, facendo riferimento anche ad altre azioni inerenti la progettazione di un test: scelta/ scrittura dei testi, elaborazione dei *prompt*; scelta dell'*input* di appoggio, ecc.

⁸⁰ Muitas vezes o termo é usado de uma maneira mais ampla, referindo-se também a outras ações inerentes à elaboração de um teste: escolha/ escrita de textos, elaboração dos *prompt*; escolha do *input* de apoio, etc.

⁸¹ Il termine *candidato* è, in verità, più ristretto, poiché riferito a coloro che si sottopongono, in particolare, a prove selettive.

⁸² O termo *candidato* é, na verdade, mais restrito, uma vez que se refere àqueles que participam especificamente de concursos seletivos.

G. CRITICITÀ

G. ASPECTOS CRÍTICOS

Illustriamo una serie di aspetti problematici che ineriscono l'uso della valutazione (*cattivi usi*), la fase di progettazione (*difetti di progettazione*), il giudizio del valutatore (*difetti del giudizio*) o l'allestimento della scala di valutazione (*difetti della scala*).

Apresentamos uma série de aspectos problemáticos relacionados ao uso da avaliação (*uso inapropriado*), à fase de elaboração (*erro de elaboração*), à nota atribuída pelo avaliador (*erro de atribuição de nota*) ou à preparação da escala de avaliação (*escala inadequada*).

Cattivi usi

Anziché strumento volto a dare valore (come vorrebbe l'etimo), la prova/ l'esercitazione viene usata come un'occasione di potere (cfr. SHOHAMY, 2001). Si considerino le seguenti circostanze:

- *si etichettano/ stigmatizzano gli studenti a seconda degli esiti raggiunti nelle prove;*⁸³

⁸³ Rosenthal e Jacobson (1977) coniarono il termine "effetto Pigmalione". In un loro studio, ad alcuni insegnanti venne detto che certi allievi presentavano un profilo alto. Il fatto non era vero, ma fu sorprendente notare che, otto mesi dopo, quegli studenti raggiunsero prestazioni migliori di quelle dei compagni. Verosimilmente gli insegnanti concorsero, in maniera inconsapevole, a modellare il profilo degli allievi in base alle aspettative che avevano su di loro, contribuendo a una "profezia che si avvera".

- *si ricorre al test come strumento di disciplina* (si pensi all'interrogazione a sorpresa o al test volto a 'calmare' gli animi di una classe riottosa);
- *mediante il test si trasmettono posizioni ideologiche e discriminatorie* (concetti connotati in senso religioso, sessuale, politico, ecc.);
- *si ricorre a parametri discrezionali* (per esempio, quesiti su questioni extra-curricolo o, nella valutazione dello scritto, considerazioni sulle idee espresse dal candidato, ecc.).

Uso inapropriado

Em vez de ser usada como instrumento para dar valor (como a etimologia alude), a prova/ a atividade é usada como um pretexto para abuso de poder (cfr. SHOHAMY, 2001). Considere as seguintes circunstâncias:

- *os estudantes são rotulados/ estigmatizados de acordo com os resultados obtidos nas provas;*⁸⁴
- *o teste é usado como uma ferramenta de disciplina* (pense na prova 'surpresa' ou no teste destinado a 'acalmar' os ânimos de uma turma rebelde);
- *posições ideológicas e discriminatórias são transmitidas através do teste* (conceitos com conotação ideológica de cunho religioso, sexual, político, etc.);

⁸⁴ Rosenthal e Jacobson (1977) cunharam o termo "efeito Pigmalão". Em um de seus estudos, alguns professores foram informados de que certos alunos tinham um alto perfil. O fato não era verdadeiro, mas foi surpreendente notar que, oito meses depois, esses alunos alcançaram um desempenho melhor do que seus colegas. Provavelmente os professores, inconscientemente, ajudaram a moldar o perfil dos alunos de acordo com sua expectativas, contribuindo para uma "profecia que se torna realidade".

- *são utilizados parâmetros subjetivos* (por exemplo, questões sobre temas extracurriculares ou, na avaliação da escrita, comentários sobre as ideias expressas pelo candidato, etc.).

Difetti di progettazione

Formulazione inadeguata degli *item*; scelta inappropriata del testo; formulazione confusa del *prompt*; vaghezza delle istruzioni; inadeguatezza dell'**input di appoggio**.⁸⁵

Erros de elaboração

Formulação inadequada dos itens; escolha inadequada do texto; formulação confusa do *prompt*; imprecisão nas instruções; inadequação do **input de apoio**.⁸⁶

Difetti del giudizio

Esistono diverse fonti di distorsione del giudizio:

- *effetto alone*: la valutazione della prestazione si intreccia con questioni di antipatia/simpatia o con indebite generalizzazioni (per esempio, in una prova di produzione orale, il fatto che il candidato sia molto accurato nell'espressione

⁸⁵ Ulteriori criticità nell'ambito delle attività di comprensione:

- *fattore riconoscimento*; la ripresa letterale tra *item* e informazione *target*;
- *predittività*; lo studente risolve un *item* sulla base delle conoscenze personali (senza attingere le informazioni dal testo di riferimento; POWERS, WILSON, 1993; KATZ, LAUTENSCHLAGER, 1994; TORRESAN, 2016);
- doppia chiave o distrattore parzialmente vero.

Per un prospetto esteso, cfr. Torresan 2019, 2020a.

⁸⁶ Outros aspectos críticos nas atividades de compreensão são:

- *fator reconhecimento*; a reprodução literal da informação *target* em parte do item;
- *previsibilidade*; o estudante resolve um item com base em conhecimentos pessoais (sem extrair informações do texto de referência; POWERS, WILSON, 1993; KATZ, LAUTENSCHLAGER, 1994; TORRESAN, 2016);
- duas opções corretas ou distrator parcialmente verdadeiro.

Para um panorama amplo, cfr. Torresan 2019, 2020a.

porta a ritenere che sia molto competente in generale);

- *effetto di prossimità*: il parere sulla prestazione di un candidato è influenzato dal giudizio *appena* espresso sulla prestazione di altri candidati (un esaminatore può, così, sopravvalutare una produzione mediocre a seguito di aver preso visione di alcune produzioni scarse);
- *tendenza alla centralizzazione*: il giudizio tende a limitarsi a un intervallo medio di punteggio;
- *effetto distribuzione forzata dei risultati*: il valutatore assume che la **distribuzione normale** valga in qualsiasi contesto; se insegnante, può ritenere che, qualsiasi sia la classe, sempre si danno poche prestazioni scarse, poche eccellenti, e molte attorno alla sufficienza (ANTIBI, 2003).

Erros de avaliação

Existem várias fontes de distorção da avaliação:

- *efeito halo*: a avaliação está interligada a questões de antipatia/ simpatia ou a generalizações indevidas (por exemplo, em uma prova de produção oral, o fato de o candidato ser acurado ao falar sugere que ele seja muito competente em geral);
- *efeito de proximidade*: a nota de um candidato é influenciada pelas notas que acabam de ser conferidas na avaliação de outros candidatos (um examinador pode, dessa forma, superestimar uma produção medíocre após ter examinado algumas produções insuficientes);

- *tendência à centralização*: a nota tende a ficar limitada em um intervalo médio de pontuação;
- *efeito distribuição forçada dos resultados*: o avaliador assume que a **distribuição normal** é válida em qualquer contexto; se professor, ele pode considerar que, qualquer que seja a turma, há sempre poucos com desempenho insuficiente, poucos com desempenho excelente e muitos com desempenho suficiente (ANTIBI, 2003).

Difetti della scala

Una **scala** può essere inadeguata nella misura in cui

- non riflette gli intervalli di competenza;
- non permette di considerare le prestazioni peggiori (*effetto pavimento* o *floor effect*);
- non permette di considerare le prestazioni migliori (*effetto tetto* o *ceiling effect*).

Escala inadequada

Uma **escala** pode ser inadequada na medida em que

- não reflete os níveis de competência;
- não permite considerar os piores desempenhos (*efeito de piso* ou *floor effect*);
- não permite considerar os melhores desempenhos (*efeito teto* ou *ceiling effect*).

Difetti nella definizione del costrutto operativo

Un'attività può presentare problemi in relazione al costrutto che la caratterizza:

- *indeterminazione del costrutto*: **attività integrate**, come il *dettato* e il *cloze* tradizionale (con una

cancellazione di una parola ogni sette), pur tornando utili in sede didattica per esercitare abilità (nei casi di dettato e *cloze*, rispettivamente ascolto e lettura), non possono essere usate in sede valutativa a tale scopo. Risulta, in effetti, complesso stabilire quali siano le **componenti** che presiedano ai rispettivi **costrutti**;⁸⁷

- *sottorappresentazione del costrutto*; la prova operativizza una parte molto ridotta del **costrutto di riferimento** (approfondimento all'appendice 5);
- *variabile esterna al costrutto*: la prestazione del candidato è fortemente influenzata (in un senso positivo o negativo) da fattori esterni al **costrutto di riferimento** (approfondimento all'appendice 6).

Erro na definição do constructo operacional

Uma atividade pode apresentar problemas em relação ao constructo:

- *indeterminação do constructo*: **atividades integradas**, como o *ditado* e o *cloze* tradicional (com uma palavra cancelada a cada sete), embora sejam úteis no ensino para exercitar habilidades (nos casos de ditado e *cloze*, respectivamente, a escuta e a leitura), não podem ser usadas na avaliação para este fim. É, de fato, complexo determinar quais

⁸⁷ Per chi voglia studiare più a fondo la questione, suggeriamo i seguenti riferimenti:

- per il dettato: Lado, 1961; Balboni, 1998, 2008.
- per il *cloze*: Carroll *et al.*, 1959; Alderson, 1979; Hansen, Stansfield, 1981, 1982, 1983; Porter, 1983; Chapelle, Abraham, 1990; Chiari, 2002.

são os **componentes** que formam os respectivos **constructos**;⁸⁸

- *sub-representação do constructo*; a prova testa uma pequena parte do **constructo de referência** (ler mais no [apêndice 5](#));
- *variável externa ao constructo*: o desempenho do candidato é fortemente influenciado (no sentido positivo ou negativo) por fatores externos ao **constructo de referência** (ler mais no [apêndice 6](#)).

⁸⁸ Para quem quiser aprofundar o assunto, sugerimos as seguintes referências:

- para o ditado: Lado, 1961; Balboni, 1998, 2008.
- para o *cloze*: Carroll *et al.*, 1959; Alderson, 1979; Hansen, Stansfield, 1981, 1982, 1983; Porter, 1983; Chapelle, Abraham, 1990; Chiari, 2002.

H. MISURAZIONI, RAPPRESENTAZIONI E VALORI

H. MEDIÇÕES, REPRESENTAÇÕES E VALORES

Raccogliamo in questa sede tutti quegli strumenti di cui il valutatore può disporre per rappresentare e analizzare i risultati di un test. Per chi volesse documentarsi ulteriormente, consigliamo la lettura di McNamara, 1996, di Carr, 2011 e di Green, 2013.

Reunimos a seguir todas as ferramentas que o avaliador pode utilizar para representar e analisar os resultados de um teste. Para aqueles que querem mais informações, recomendamos a leitura de McNamara, 1996, de Carr, 2011 e de Green, 2013.

Analisi dei distrattori (*Distractor Analysis*)

Studio del comportamento dei **distrattori** (cfr. **indice di attrattività**).

Análise dos distratores (*Distractor Analysis*)

Estudo do comportamento dos **distratores** (cfr. **parâmetro de atração do distrator**).

Analisi fattoriale Misurazione volta a definire raggruppamenti di fattori (o **componenti**), che si configurano nella forma di un **costrutto**.

Análise fatorial Medição para definir agrupamentos de fatores (ou **componentes**), que definem um **constructo**.

Analisi degli *item* (*Item Analysis*)

Studio del comportamento degli *item*, mediante valori della statistica descrittiva (**indice di facilità, indice di discriminazione, indice di affidabilità Alfa di Cronbach, deviazione standard**).⁸⁹

Análise dos itens (*Item Analysis*)

Estudo do comportamento dos itens, utilizando valores da estatística descrittiva (**índice de facilidade, índice de discriminação, índice de confiabilidade Alfa de Cronbach, desvio padrão**).⁹⁰

Analisi Rasch

Il modello Rasch consente di linearizzare (disporre su un unico parametro) due⁹¹ o più **variabili**,⁹² le quali abbiano un carattere unidimensionale, ovvero facciano riferimento a un medesimo **costrutto** (cfr. BOND, FOX, 2007).

La linearizzazione viene operata mediante una **scala** a zero centrale, con valori positivi e negativi (simile alla scala Celsius per le temperature). I valori, chiamati **logit** (*log-odd units*), sono dei logaritmi.

L'analisi di due variabili viene applicata ai risultati di **attività strutturate** (per esempio test di comprensione chiusi). L'analisi verifica

- se gli *item* restituiscono una mappa adeguata dell'abilità dei candidati;

⁸⁹ Lo studio può essere condotto anche attraverso **l'Analisi Rasch**.

⁹⁰ O estudo também pode ser realizado através da **Análise Rasch**.

⁹¹ Il *software* di riferimento è Winsteps ©.

⁹² Il *software* di riferimento è Facets ©.

- se il campione di candidati permette di tracciare una mappa adeguata del grado di difficoltà degli *item* (GREEN 2013: 150)⁹³.

L'analisi di più di due variabili viene applicata ad **attività aperte**, quali test di produzione scritta/orale. Mediante la scala comune si mettono in relazione

- il grado di difficoltà del compito,
- il grado di severità dell'esaminatore,
- l'abilità del candidato.

In sostanza, si accerta quanto l'esaminatore sia stato severo e, al tempo stesso, se il compito sia stato adeguatamente calibrato, in rapporto alla competenza dei candidati.

Análise Rasch

O modelo Rasch permite linearizar (arranjar em um único parâmetro) duas⁹⁴ ou mais **variáveis**,⁹⁵ que tenham um caráter unidimensional, ou seja, que se refiram a um mesmo **constructo** (cfr. BOND, FOX, 2007).

A linearização é realizada através de uma **escala** que tem o zero como referência central, com valores positivos e negativos (semelhante à escala Celsius para temperatura). Os valores, chamados **logits** (*log-odd units*), são logaritmos.

⁹³ In questo caso è possibile ricorrere a una *tabella di conversione* (<www.winsteps.com/winman/whatisalogit.htm>) per trasformare la differenza di *logit* tra le due variabili in termini di probabilità. Un candidato di 1 *logit*, per esempio, ha il 73% di probabilità di risolvere con successo un *item* di 0 *logit*; un candidato di -1 *logit* ha il 27% di probabilità di risolvere con successo un *item* di 0 *logit*. Nell'*analisi di più di due variabili* i calcoli sono estremamente più complessi, e non si dispone di una tabella di conversione.

⁹⁴ O *software* de referência é Winsteps ©.

⁹⁵ O *software* de referência é Facets ©.

A análise de duas variáveis é aplicada aos resultados de **atividades com gabarito**. A análise verifica

- se os itens fornecem um mapa apropriado da habilidade dos candidatos;
- se a amostra de candidatos permite desenhar um mapa apropriado do grau de dificuldade dos itens (GREEN, 2013: 150)⁹⁶.

A análise de mais de duas variáveis é aplicada em **atividades com resposta livre**, tais como testes de produção escrita/oral. Por meio da escala comum se relacionam

- o grau de dificuldade da tarefa,
- o grau de severidade do examinador,
- a habilidade do candidato.

Basicamente, se verifica quanto o examinador foi severo e, ao mesmo tempo, se a tarefa foi devidamente adequada em relação à competência dos candidatos.

Affidabilità (índice)

L'índice di affidabilità maggiormente diffuso nell'ambito della statistica classica è l'*Alfa di Cronbach*. L'índice ci restituisce il grado di **omogeneità** tra gli *item* ed è calcolato sulla base delle **varianze** (è dato dal rapporto tra la varianza data dai singoli *item* e la varianza

⁹⁶ Neste caso é possível usar uma *tabela de conversão* (<www.winsteps.com/winman/whatisalogit.htm>) para transformar a diferença de *logit* entre as duas variáveis em termos de probabilidade. Um candidato de 1 *logit*, por exemplo, tem uma probabilidade de 73% de resolver com sucesso um item de 0 *logit*; um candidato de -1 *logit* tem uma probabilidade de 27% de resolver com sucesso um item de 0 *logit*. Na *análise de mais de duas variáveis*, os cálculos são extremamente mais complexos e não existe uma tabela de conversão.

complessiva dell'esercitazione [ovvero dell'intero blocco di *item*]). Il valore è compreso nell'intervallo ± 1 . Valori negativi indicano il fatto che alcuni *item* 'viaggiano' in senso contrario rispetto ad altri (occorre indagare, in tal caso, se una *variabile esterna* al **costrutto** non stia compromettendo la **validità** del test).

Il valore atteso è

> .70	GREEN, 2013
-------	-------------

Concorrono all'affidabilità di un esercizio

- la **discriminatività** degli *item* (maggiore è il grado di discriminatività, maggiore è l'affidabilità);
- la lunghezza dell'esercizio (il quale deve contemplare almeno 10 *item*);⁹⁷
- la **calibrazione** degli *item*;
- la chiarezza delle istruzioni;
- l'accuratezza delle **chiavi**;
- il livello di **familiarità** del candidato con il tipo di attività;
- il grado di indipendenza dei segmenti di conoscenza/competenza misurati dai singoli *item* (per esempio, il fatto che due *item* di una prova di comprensione si riferiscano a **informazioni target** distinte, e non alla stessa informazione).⁹⁸

⁹⁷ Per esercizi con un numero di *item* inferiore a 10, vale la pena operare un'analisi della **correlazione** interna tra gli *item* (*Inter-item Correlation*), anziché ricorrere al suddetto indice (GREEN, 2013: 39-40).

⁹⁸ Scrivono Schedl, Malloy (2004: 8): "Poiché ogni *item* contribuisce alla possibilità di formulare delle inferenze in merito all'abilità di un candidato, è importante che gli *item* siano indipendenti

Vale la pena ricordare che la rimozione di un *item* da una **batteria** comporta una variazione dell'affidabilità degli altri *item*. È importante perciò avviare, a seguito di ogni modifica all'assetto di una batteria, un nuovo calcolo dell'indice.

Confiabilidade (índice)

O índice de confiabilidade mais comum na estatística clássica é o *Alfa de Cronbach*. O índice nos informa o grau de **homogeneidade** entre os itens e é calculado com base nas **variâncias** (é determinado pela divisão entre a variância dada pelos itens individuais e a variância global da atividade [ou seja, de todo o conjunto de itens]). O valor deve estar contido no intervalo de ± 1 . Valores negativos apontam para o fato de que alguns itens 'se movem' na direção oposta em relação a outros (neste caso, é necessário investigar se uma *variável externa* ao **constructo** não está afetando a **validade** do teste).

O valor esperado è

>.70	GREEN, 2013
------	-------------

Contribuem para a confiabilidade de um exercício

- a **discriminação** dos itens (quanto maior for o grau de discriminação, maior será a confiabilidade);

[tra loro], nel senso che ciascuno di essi deve riferirsi a questioni distinte. La mancanza di indipendenza riduce l'affidabilità complessiva di un test?.

- a duração do exercício (que deve incluir pelo menos 10 itens);⁹⁹
- a **adequação** dos itens;
- a clareza nas instruções;
- a exatidão do **gabarito**;
- o nível de **familiaridade** do candidato com o tipo de atividade;
- o grau de independência dos segmentos de conhecimento/competência medidos por cada item (por exemplo, o fato de que, em uma prova de compreensão, dois itens se referem a **informações target** distintas, e não à mesma informação).¹⁰⁰

Vale lembrar que a remoção de um item de um **banco de itens** causa uma variação na confiabilidade dos outros itens. Por conseguinte, é importante iniciar um novo cálculo de índice após qualquer alteração da estrutura do banco de itens.

Asimetria (coeficiente)

Definisce il disallineamento dei dati rispetto alla **media**.

Il coefficiente può avere valori positivi o negativi:

⁹⁹ Para exercícios com um número de itens inferior a 10, vale a pena fazer uma análise da **correlação** interna entre os itens (*Inter-item Correlation*), em vez de usar o índice acima mencionado (GREEN, 2013: 39-40).

¹⁰⁰ Schedl, Malloy escrevem (2004: 8): “Uma vez que cada item contribui para fazer inferências sobre a habilidade de um candidato, é importante que os itens sejam independentes [um do outro], o que significa que cada item deve se referir a questões distintas. A falta de independência reduz a confiabilidade geral de um teste”.

- valori positivi riflettono una prova relativamente difficile, con un numero maggiore di risultati bassi (si parla anche di *distribuzione asimmetrica a destra*, posto che la coda più lunga della curva si trova a destra del valore massimo centrale, cioè della **moda**);
- valori negativi riflettono una prova relativamente facile, con un numero maggiore di risultati alti (si parla anche di *distribuzione asimmetrica a sinistra*, posto che la coda più lunga della curva si trova a sinistra del valore massimo centrale, cioè della **moda**).

Una **distribuzione normale** ha un'asimmetria nulla.

Assimetria (coefficiente)

Define o desalinhamento dos dados em relação à **média**. O coeficiente pode ter valores positivos ou negativos:

- valores positivos refletem uma prova relativamente difícil, com um maior número de resultados ruins (também se fala de *distribuição assimétrica à direita*, uma vez que a linha mais longa da curva está localizada à direita do valor máximo central, que é a **moda**);
- valores negativos refletem uma prova relativamente fácil, com um número maior de resultados bons (também se fala de *distribuição assimétrica à esquerda*, uma vez que a linha mais longa da curva está localizada à esquerda do valor máximo central, que é a **moda**).

Uma **distribuição normal** tem assimetria zero.

Attrattività (indice)

Percentuale di consensi riscossi da un determinato **distrattore**. I valori di riferimento per *prove di competenza* e per *prove di fine ciclo* sono suggeriti qui sotto.

10		BACHMAN, 2004 ¹⁰¹
7	30	Nostra proposta ¹⁰²

Capacidade de atração do distrator (índice)

Percentual de candidatos que escolheram um **distrator** como opção correta. Os valores de referência para *provas de proficiência* e para *provas de conclusão* estão sugeridos abaixo.

10		BACHMAN, 2004 ¹⁰³
7	30	Nossa proposta ¹⁰⁴

Comportamento differenziato degli *item* (DIF, *Differential Item Functioning*)

Studio del funzionamento di uno o più *item* in sottogruppi significativi di una certa popolazione (sottogruppi distinguibili per età, genere, grado di istruzione, ecc.).

Se l'*item* assume un comportamento significativamente diverso da sottogruppo a sottogruppo, pur considerando che i sottogruppi abbiano una pari abilità,

¹⁰¹ Bachman specifica solo il valore minimo.

¹⁰² Valori inferiori alla soglia minima (<7%) ci informano del fatto che il distrattore è inefficace; valori superiori (>30%) ci informano di una possibile doppia chiave.

¹⁰³ Bachman especifica apenas o valor mínimo.

¹⁰⁴ Valores abaixo do limiar mínimo (<7%) nos informam que o distrator é ineficaz; valores mais altos (>30%) nos informam sobre uma possível segunda opção correta.

dev'essere riformulato o sostituito, posto che una **variabile esterna** inficia la **validità** dell'esercitazione (si immagini, per esempio, che i risultati di un test di comprensione da parte di candidati di pari abilità siano significativamente diversi, e che la disparità sia imputabile al diverso grado di **familiarità** con il tipo di attività; cfr. SHOHAMY, 1984).¹⁰⁵

Comportamento diferenciado dos itens (DIF, *Differential Item Functioning*)

Estudo do funcionamento de um ou mais itens em subgrupos significativos de uma determinada população (subgrupos distinguíveis por idade, sexo, escolaridade, etc.). Se o item assumir um comportamento significativamente diferente de subgrupo para subgrupo, mesmo considerando que os subgrupos têm capacidade igual, deve ser reformulado ou substituído, já que uma **variável externa** prejudica a **validade** da atividade (imagine, por exemplo, que os resultados de um teste de compreensão são significativamente diferentes mesmo tendo candidatos de igual capacidade, e que tal disparidade é atribuível ao diferente grau de **familiaridade** com o tipo de atividade; cfr. SHOHAMY, 1984).¹⁰⁶

¹⁰⁵ Per approfondimenti, cfr. Chen, Henning, 1985; Kunnan, 1990; Sasaki, 1991; Shohamy, Inbar, 1991; Ryan, Bachman, 1992; Holland, Wainer, 1993; Camilli, Shepard, 1994; Kim, 2001; Pae, 2004; Geranpayeh, Kunnan, 2007; Roever, 2007; Geranpayeh, 2008; Karami, Salmani Nodoushan, 2011.

¹⁰⁶ Para maiores informações, cfr. Chen, Henning, 1985; Kunnan, 1990; Sasaki, 1991; Shohamy, Inbar, 1991; Ryan, Bachman, 1992; Holland, Wainer, 1993; Camilli, Shepard, 1994; Kim, 2001; Pae, 2004; Geranpayeh, Kunnan, 2007; Roever, 2007; Geranpayeh, 2008; Karami, Salmani Nodoushan, 2011.

Correlazione (indice)

Esprime il vincolo di co-occorrenza (non necessariamente causale) tra due **variabili**. Il calcolo di una correlazione si basa sulla **covarianza**. Il valore oscilla tra ± 1 ;

- valori positivi indicano una *correlazione positiva*;
- valori negativi indicano una *correlazione negativa*;
- valori prossimi allo zero indicano un'*assenza di correlazione*.

Per quanto concerne la *forza di una correlazione*, si confrontino le soglie riportate qui sotto.

Debole	Moderata	Forte	Fonte
<.030	0.30-0.70	> 0.70	POZZOLO, 2020 ¹⁰⁷

Il calcolo del coefficiente di correlazione varia in funzione delle caratteristiche della **distribuzione**:

- si calcola l'*indice di correlazione Pearson* nel caso di test parametrici (con una **distribuzione** che rientra all'interno di una soglia di **normalità**);
- si calcola l'*indice di correlazione Spearman* nel caso di test non parametrici (con una **distribuzione** lontana dalla **normalità**).¹⁰⁸

¹⁰⁷ Cfr. anche Dancey, Reidy, 2007.

¹⁰⁸ La soglia della normalità della distribuzione è calcolata mediante la divisione dei **coefficienti di asimmetria** e di **curtosi** per i rispettivi **errori standard di misurazione**. Se l'esito della divisione, in entrambi i casi, è contenuto all'interno del valore ± 2 , si adotta l'*indice di Pearson*; se eccede questo intervallo si adotta l'*indice di Spearman*.

A restituire l'idea della forza di una correlazione è anche il *coefficiente di determinazione*, definito dal quadrato del *coefficiente di correlazione*, e direttamente convertibile in percentuale. A un coefficiente di correlazione di 0.7, per esempio, corrisponde un *coefficiente di determinazione* pari a 49: quest'ultimo indica una sovrapposizione del 49% (7x7) tra le due **variabili**.

Correlação (índice)

Expressa a relação de coocorrência (não necessariamente causal) entre duas **variáveis**. O cálculo de uma correlação é baseado na **covariância**. O valor oscila entre ± 1 ;

- valores positivos indicam uma *correlação positiva*;
- valores negativos indicam uma *correlação negativa*;
- valores próximos de zero indicam uma *ausência de correlação*.

Quanto à *força de uma correlação*, comparar os limiares indicados abaixo.

Fraca	Moderada	Forte	Fonte
<.030	0.30-0.70	> 0.70	POZZOLO, 2020 ¹⁰⁹

O cálculo do coeficiente de correlação varia de acordo com as características da **distribuição**:

¹⁰⁹ Cfr. também Dancey, Reidy, 2007.

- calcula-se o *índice de correlação Pearson* no caso de testes paramétricos (com uma **distribuição** dentro de um limiar de **normalidade**);
- calcula-se o *índice de correlação Spearman* no caso de testes não paramétricos (com uma **distribuição** longe da **normalidade**).¹¹⁰

Para ter uma ideia da força de uma correlação também vale o *coeficiente de determinação*, definido pelo quadrado do *coeficiente de correlação*, e diretamente convertível em percentual. Um coeficiente de correlação de 0.7, por exemplo, corresponde a um *coeficiente de determinação* de 49: este último indica uma sobreposição de 49% (7x7) entre as duas **variáveis**.

Covarianza (índice)

Indica il grado di accordo della variazione di due variabili attorno alle rispettive **medie**.

Covariância (índice)

Indica o grau de concordância da variação de duas variáveis em torno das respectivas **médias**.

Curtosi (coeficiente)

Definisce l'approssimazione dei dati attorno alla **media**.

Può avere valori positivi o negativi:

¹¹⁰ O limiar da normalidade da distribuição é calculado dividindo os **coeficientes de assimetria** e de **curtose** pelos seus respectivos **erros de medida padrão**. Se o resultado das divisões, em ambos os casos, está contido dentro do valor ± 2 , o *índice de Pearson* é adotado; se excede este intervalo, o *índice de Spearman* é adotado.

- valori positivi riflettono un comportamento omogeneo dei candidati, con risultati che si aggirano attorno alla media (“distribuzione leptocurtica”);
- valori negativi riflettono una curva piatta, con una notevole variazione dei risultati (“distribuzione platicurtica”);
- con un valore pari a zero, abbiamo una distribuzione normale (cfr. **normalità**).

Curtose (coefficiente)

Define a aproximação dos dados em torno da **média**.
 Pode ter valores positivos ou negativos:

- valores positivos refletem um comportamento homogêneo dos candidatos, com resultados que estão em torno da média (“distribuição leptocúrtica”);
- valores negativos refletem uma curva achatada, com uma variação significativa dos resultados (“distribuição platicúrtica”);
- com um valor igual a zero, temos uma distribuição normal (cfr. **normalidade**).

Discriminazione (indice e *correlazione Punto biseriale*)

Un *item* **discrimina** nella misura in cui distingue i candidati più abili da quelli meno abili. Per calcolare l'indice occorre suddividere la popolazione in tre gruppi:

- candidati con fascia di punteggio superiore;
- candidati con fascia di punteggio medio;

- candidati con fascia di punteggio inferiore.

È importante che il primo e il terzo gruppo siano costituiti dallo stesso numero di individui (ciascuno pari al 25-27% del totale; cfr. KUNNAN, CARR, 2012). La discriminatività del singolo *item* è calcolata mediante la formula

$$(CS-CI)/nS$$

dove

- CS è il numero di apprendenti del gruppo di punteggio superiore che hanno risposto correttamente all'*item*;
- CI è il numero di apprendenti del gruppo di punteggio inferiore che hanno risposto correttamente all'*item*;
- nS è il numero di apprendenti che fanno parte del gruppo di punteggio superiore.

In genere, all'*indice di discriminatività* si tende comunque a preferire la **correlazione Punto biseriale** (o più semplicemente *Punto biseriale*, r_{p-bis}), specie se si dispone di un campione relativamente ampio (>35; cfr. KUNNAN, CARR, 2012). Il *Punto biseriale* tiene conto, infatti, del comportamento di tutti i candidati e non solo di quelli delle fasce estreme. Questo valore mette in relazione il comportamento del singolo *item* con quello degli altri *item*.¹¹¹

¹¹¹ Nel pacchetto statistico SPSS è chiamato *Corrected item-total correlation*: si tratta di una correlazione tra il singolo elemento e l'intera batteria.

Entrambi gli indici (*indice di discriminatività e correlazione Punto biseriale*) oscillano all'interno dell'intervallo ± 1 . Il valore atteso è

> .30	SCHEDL, MALLOY, 2014 GRABOWSKI, DAKIN, 2014
-------	--

A valori negativi corrisponde una discriminazione negativa (i più bravi sbagliano, i meno bravi indovino). Come si diceva riguardo all'**affidabilità**, l'eventuale rimozione di un *item* da una batteria, una volta appuratane la bassa *discriminatività*, comporta una variazione dei valori degli altri *item*. È importante perciò avviare, a seguito di una eventuale modifica al test, un nuovo calcolo dell'indice.

Discriminação de elementos (*índice e coeficiente de correlação Ponto-biserial*)

Um item **discrimina** na medida em que distingue os candidatos mais capazes daqueles menos capazes. Para calcular o índice, a população deve ser dividida em três grupos:

- candidatos na faixa de pontuação alta;
- candidatos na faixa de pontuação média;
- candidatos na faixa de pontuação baixa.

É importante que o primeiro e o terceiro grupos sejam compostos pelo mesmo número de indivíduos (cada um igual a 25-27% do total; cfr. KUNNAN, CARR, 2012). A discriminação de um único item é calculada através da fórmula

$$(CS-CI)/nS$$

onde

- CS é o número de alunos do grupo com pontuação alta que respondeu corretamente ao item;
- CI é o número de alunos do grupo com pontuação baixa que respondeu corretamente ao item;
- nS é o número total de alunos que pertence ao grupo com pontuação alta.

Em geral, tende-se a preferir a **correlação Ponto-bisserial** (ou, simplesmente, *Ponto bisserial*, r_{p-bis}) ao *índice de discriminação*, sobretudo com uma amostragem relativamente ampla (>35; cfr. KUNNAN, CARR, 2012). O *Ponto bisserial* leva em conta o comportamento de todos os candidatos, não se restringe apenas aos candidatos com os melhores ou piores resultados. Este valor relaciona o comportamento de um único item com o de outros itens.¹¹²

Ambos os índices (*índice de discriminação* e *correlação Ponto-bisserial*) oscilam entre ± 1 . O valor esperado é

> .30	SCHEDL, MALLOY, 2014 GRABOWSKI, DAKIN, 2014
-------	--

¹¹² Na plataforma de software estatístico SPSS é chamado *Corrected item-total correlation*: trata-se de uma correlação entre um único elemento e todo o banco de itens.

Os valores negativos correspondem a uma discriminação negativa (os melhores erram, os mais fracos acertam).

Como dissemos sobre a **confiabilidade**, a possível remoção de um item de um banco de itens, uma vez constatada a baixa *discriminação*, causa uma variação nos valores dos outros itens. Por conseguinte, é importante iniciar um novo cálculo do índice após qualquer alteração no teste.

Devianza Somma del quadrato delle distanze dei singoli dati rispetto alla **media** (Σd^2).

Desvio Soma do quadrado das distâncias de cada dado em relação à **média** (Σd^2).

Deviazione standard (o *scarto quadratico medio*)

Radice quadrata (con segno positivo) della **varianza**. Vale a indicare il grado di **distribuzione** dei risultati (maggiore è il grado di distribuzione, maggiore è la varietà).

Desvio padrão Raiz quadrada (com sinal positivo) da **variância**. Serve para indicar o grau de **distribuição** dos resultados (quanto maior o grau de distribuição, maior a variedade).

Distribuzione Frequenza dei risultati all'interno della **scala di valutazione**. Viene definita mediante il calcolo della **deviazione standard**.

Distribuição Frequência dos resultados dentro da **escala de avaliação**. É definida pelo cálculo do **desvio padrão**.

Errore di misurazione

Indica il grado di approssimazione della rilevazione. È sensibile al numero di dati osservati: maggiore è il numero, minore è l'errore.

È ottenuto mediante il prodotto della **deviazione standard** con la radice quadrata di 1 meno **l'indice di affidabilità**.

Se l'errore è 0.2 e abbiamo un valore osservato di 1, il “vero” valore è compreso tra

- 0.80 e 1.20 (cioè 1 ± 0.20), con un 68% di probabilità;
- 0.60 e 1.40 (cioè $1 \pm 0.20 \times 2$), con un 95% di probabilità.¹¹³

Erro de medida Indica o grau de imprecisão da medição. É sensível ao número de dados observados: quanto maior o número, menor o erro.

É obtido pelo produto do **desvio padrão** com a raiz quadrada de 1 menos o **índice de confiabilidade**.

Se o erro é 0.2 e temos um valor observado de 1, o valor “verdadeiro” está entre

- 0.80 e 1.20 (isto é, 1 ± 0.20), com uma probabilidade de 68%;

¹¹³ Per il significato di queste percentuali si veda la voce **normalità**.

- 0.60 e 1.40 (isto é, $1 \pm 0.20 \times 2$), com uma probabilidade de 95%.¹¹⁴

Facilità (indice; *facility value*, FV)

Si riferisce al numero di candidati che risolvono con esito il singolo *item*. Il valore è espresso mediante una percentuale (es. 50%) o un numero (50, 0.50 o .50).

In merito ai valori attesi in *prove di competenza* le indicazioni presenti nella letteratura di riferimento divergono.

Min.	Max	Fonte
30	70	ALLEN, YEN, 1978 BACHMAN, 2004
20	80	GREEN, 2013
25	75	BRITISH COUNCIL, 2020
25	80	ALTE, 2011

Valori inferiori rispetto alla soglia minima indicano che l'*item* è sovracalibrato (troppo difficile); valori superiori rispetto alla soglia massima indicano che l'*item* è sottocalibrato (troppo facile).

Per *prove di fine ciclo*, che presuppongono una preparazione continua dei candidati, suggeriamo il seguente intervallo

Min.	Max	Fonte
35	85	nostra proposta

¹¹⁴ Para saber o significado desses percentuais, consulte o item **normalidade**.

Per *prove di profitto*, abbiamo i seguenti valori

Min.	Max	Fonte
50	100	BRITISH COUNCIL, 2020

In sostanza, in *prove di profitto* non ci si preoccupa della soglia massima (100 è anzi il valore atteso): se l'insegnante ha insegnato bene e gli studenti si sono impegnati, ci si aspetta che l'intera classe superi il test.

Índice de facilidade (*facility value, FV*)

Refere-se ao número de candidatos que resolvem com êxito um item. O valor é expresso por um percentual (ex. 50%) ou por um número (50, 0.50 ou .50).

Em relação ao valores previstos nas *provas de proficiência*, as indicações na literatura de referência divergem.

Min.	Max	Fonte
30	70	ALLEN, YEN, 1978 BACHMAN, 2004
20	80	GREEN, 2013
25	75	BRITISH COUNCIL, 2020
25	80	ALTE, 2011

Valores inferiores ao limite mínimo indicam que o item está muito difícil; valores superiores ao limite máximo indicam que o item está muito fácil.

Para *provas de conclusão*, que requerem a preparação contínua dos candidatos, sugerimos o seguinte intervalo

Min.	Max	Fonte
35	85	nostra proposta

Para *provas de conhecimento*, temos os seguintes valores

Min.	Max	Fonte
50	100	BRITISH COUNCIL, 2020

Basicamente, nas *provas de conhecimento* não nos preocupamos com o limite máximo (100 é o valor esperado): se o professor ensinou bem e os estudantes se comprometeram, esperamos que a turma inteira passe no teste.

Grafico a barre Indica la frequenza di variabili di diverso tipo. Illustra l'esito di **indagini qualitative**.

Gráfico de barras Indica a frequência de variáveis de diferentes tipos. Apresenta o resultado de **pesquisas qualitativas**.

Grafico a dispersione Evidenzia la relazione tra due **variabili dipendenti** (ad esempio, tra i punteggi raggiunti da uno stesso campione di candidati in due attività distinte), nonché la presenza di **outliers**. La relazione tra le variabili può essere

- *positiva* (al crescere dell'una, pure l'altra cresce);
- *negativa* (al crescere dell'una, l'altra diminuisce);
- *casuale* (non si evidenzia alcun *pattern* lineare).

Gráfico de dispersão

Destaca a relação entre duas **variáveis dependentes** (por exemplo, entre a pontuação alcançada pela mesma amostra de candidatos em duas atividades distintas), bem como a presença de **outliers**. A relação entre as variáveis pode ser

- *positiva* (à medida que uma cresce, a outra cresce também);
- *negativa* (à medida que uma cresce, a outra diminui);
- *aleatória* (não se destaca nenhum padrão linear).

Indagini qualitative

Permettono di apprezzare/ descrivere aspetti che ineriscono il vissuto del *test taker* (es. abitudini di lettura/ di ascolto; **strategie metacognitive**; preferenze verso i *task*; stato d'ansia durante lo svolgimento di una esercitazione; ecc.).¹¹⁵

Possono essere realizzate mediante

- strumenti aperti (es. questionari, interviste);
- strumenti chiusi (es. scala Likert);
- strumenti semistrutturati (es. lista di frasi da completare);

¹¹⁵ Scrive Dörnyei (2007: 38): “La ricerca qualitativa ha a che fare con le opinioni, le esperienze e le emozioni degli individui; [...] lo scopo è esplorare il punto di vista dei partecipanti in merito all’esperienza oggetto di indagine. [...] Questo tipo di ricerca coglie il comportamento umano come una realtà basata sui significati che gli individui attribuiscono [...] alle situazioni (PUNCH, 2005) [...]. Coloro i quali svolgono indagini qualitative si sforzano, dunque, di cogliere i fenomeni sociali con gli occhi di coloro che li vivono sulla propria pelle [...]”.

Considerando la frequenza delle risposte, i risultati possono essere oggetti a quantificazione.

Pesquisas qualitativas

Peritem valorizar/ descrever aspectos que se relacionam com a experiência do examinando (ex. hábitos de leitura/ de escuta; **estratégias metacognitivas**; preferência por tarefas; estado de ansiedade durante a realização de uma atividade; etc.).¹¹⁶

Podem ser realizadas através de

- instrumentos abertos (ex. questionários, entrevistas);
- instrumentos fechados (ex. escala Likert);
- instrumentos semiestruturados (ex. lista de frases para completar);

Considerando a frequência das respostas, os resultados podem ser quantificados.

Indici di conformità (*Fit Statistics*)

Restituiscono il grado di conformità dei dati al modello probabilistico dell'**Analisi Rasch**.

In riferimento a un'analisi condotta su due variabili, quali

- l'abilità di comprensione,

¹¹⁶ Dörnyei (2007: 38) escreve: “A pesquisa qualitativa tem a ver com opiniões, experiências e emoções dos indivíduos; [...] o objetivo é explorar o ponto de vista dos participantes a respeito da experiência que está sendo pesquisada. [...] Este tipo de pesquisa pressupõe que o comportamento humano seja uma realidade baseada nos significados que os indivíduos atribuem [...] às situações (PUNCH, 2005) [...]. Aqueles que realizam pesquisas qualitativas, portanto, se esforçam para compreender os fenômenos sociais com os olhos de quem os vive em sua própria pele [...]”.

- la difficoltà dell'*item*,

il modello prefigura che

- un candidato abile abbia un'elevata probabilità di risolvere un *item* facile;
- un candidato poco abile abbia un'alta probabilità di rispondere in maniera incorretta a un *item* difficile;
- un candidato abbia una percentuale del 50% di rispondere correttamente a un *item* calibrato sul suo livello di competenza.

Per analisi condotte su più variabili, quali

- l'abilità di produzione orale/scritta,
- il grado di severità dell'esaminatore,
- la difficoltà del compito,

il modello opera previsioni analoghe:

- l'esaminatore severo molto probabilmente assegna dei punteggi relativamente bassi alle *performance* scarse;
- l'esaminatore generoso molto probabilmente assegna dei punteggi alti a *performance* eccellenti;
- di fronte a *performance* allineate al grado di severità dell'esaminatore (es. una *performance* di *media* qualità, in rapporto a un esaminatore di *media* severità), abbiamo una certa indecidibilità (la probabilità si riduce al 50%).

Gli indici di conformità sono due:

- *Outfit* (*outlier-sensitive infit statistics*);
- *Infit* (*information weighted infit statistics*).

Entrambi si basano sui *residui*, ovvero gli scarti tra i dati attesi dal modello¹¹⁷ e i dati osservati. Il primo, l'indice *Outfit*, è sensibile ai valori *occasional* che divergono significativamente dalla norma (**outliers**). L'indice *Infit* è invece un valore ponderato (assegna un peso minore alle valutazioni estreme); esso vale a rappresentare problemi sistematici (indica, cioè, se e quanto il *pattern* di giudizio dell'esaminatore presenta criticità; ECKES, 2005, 2015²).

I valori di riferimento degli indici variano in funzione dell'entità della prova (se la prova è di alto impatto, i margini sono più stringenti). In linea di massima ci si può attenere all'intervallo **0.50-1.50** (LINACRE, 2008):

- a valori **<0.50** corrisponde una predittività maggiore del previsto (è il caso di *item* interdipendenti o di valutazioni della *performance* contraddistinte da una *tendenza centrale* o da un *effetto alone*);
- a valori **>1.50** corrisponde un'imprevedibilità superiore al previsto (il candidato risponde a caso oppure l'esaminatore non è affatto coerente).

Índices de conformidade (*Fit Statistics*)

Indicam o grau de conformidade dos dados ao modelo de probabilidade da **Análise Rasch**.

No que se refere a uma análise de duas variáveis, tais como

¹¹⁷ Per esempio, in riferimento al candidato x nello svolgimento del compito y, il modello prevede che l'esaminatore z formuli un certo giudizio, considerati i giudizi espressi dallo stesso esaminatore (z) riguardo allo svolgimento di altri compiti da parte dello stesso candidato (x), nonché i giudizi espressi dallo stesso esaminatore (z) su altri candidati.

- habilidade de compreensão,
- dificuldade de um item,

o modelo prevê que

- um candidato apto tem uma elevada probabilidade de resolver um item fácil;
- um candidato pouco apto tem uma elevada probabilidade de responder incorretamente a um item difícil;
- um candidato tem uma percentagem de 50% de resposta correta para um item adequado ao seu nível de competência.

Para análises de mais variáveis, tais como

- habilidade de produção oral/escrita,
- grau de severidade do examinador,
- dificuldade da tarefa,

o modelo faz previsões análogas:

- o examinador severo atribui, muito provavelmente, notas relativamente baixas a performances fracas;
- o examinador generoso atribui, muito provavelmente, notas altas a performances excelentes;
- diante de performances alinhadas ao grau de severidade do examinador (ex. uma performance de qualidade *média*, em relação a um examinador de severidade *média*), temos uma certa indecibilidade (a probabilidade é reduzida em 50%).

Existem dois índices de conformidade:

- *Outfit* (*outlier-sensitive infit statistics*);
- *Infit* (*information weighted infit statistics*).

Ambos se baseiam em *resíduos*, ou seja, nas diferenças entre os dados previstos pelo modelo¹¹⁸ e os dados observados. O primeiro, o índice *Outfit*, é suscetível a valores *ocasionais* que divergem significativamente da norma (**outliers**). O índice *Infit*, por outro lado, é um valor ponderado (atribui menor peso às avaliações ocasionais que divergem da norma); é válido para representar problemas sistemáticos (isto é, indica se e quanto o padrão de avaliação do examinador apresenta aspectos críticos; ECKES, 2005, 2015²).

Os valores de referência dos índices variam de acordo com o alcance da prova (se a prova é de alto impacto, os intervalos são mais estreitos). De um modo geral, deve ser mantido o intervalo de **0.50-1.50** (LINACRE, 2008):

- valores **<0.50** correspondem a uma previsibilidade maior do que o esperado (é o caso de itens interdependentes, ou de avaliações da performance marcadas por uma *tendência central* ou um *efeito halo*);
- valores **>1.50** correspondem a uma imprevisibilidade maior do que o esperado

¹¹⁸ Por exemplo, no que se refere ao candidato x ao executar a tarefa y , o modelo prevê que o examinador z dê uma determinada nota, levando em conta as notas atribuídas pelo próprio examinador (z) em relação à execução de outras tarefas pelo mesmo candidato (x), além das notas atribuídas por esse mesmo examinador (z) a outros candidatos.

(candidato risponde ao acaso, ou o examinador não é, de todo, coerente).

Intervallo (o *campo di variazione*)

Differenza tra il valore massimo e il valore minimo dei risultati; maggiore è l'intervallo, maggiore è la **distribuzione** (e quindi maggiori sono la **deviazione standard** e il grado positivo di **curtosi**).

Intervalo

Diferença entre o valor máximo e valor mínimo dos resultados; quanto maior o intervalo, maior a **distribuição** (e, portanto, maiores também o **desvio padrão** e o grau positivo de **curtose**).

Istogramma

Restituisce le frequenze di una serie di dati relativi a una **variabile continua**.

Alcuni programmi di elaborazione statistica (es. SPSS) consentono di evidenziare il profilo della curva di distribuzione.

Histograma

Informa sobre as frequências de uma série de dados relacionados a uma **variável contínua**.

Alguns programas de análise estatística (ex. SPSS) permitem destacar a curva de distribuição.

Logit

Unità di misura della scala dell'**Analisi Rasch**.

Logit

Unidade de medida da escala de **Análise Rasch**.

Media

Somma aritmetica dei risultati divisa per il numero di occorrenze (i.e. di candidati).

Média	Soma aritmética dos resultados dividida pelo número de ocorrências (i.e. de candidatos).
Mediana	Valore che si colloca a metà di una stringa di dati. ¹¹⁹
Mediana	Valor que está localizado no meio de uma sequência de dados. ¹²⁰
Moda	Valore che ricorre maggiormente in una serie di dati. ¹²¹
Moda	Valor que é aplicado mais comumente em uma série de dados. ¹²²
Normalità	È una condizione specifica di distribuzione , in cui media , mediana e moda equivalgono. Abbiamo in tal caso una perfetta simmetria dei dati (con convergenza, appunto, verso la parte centrale della curva). Maggiore è il numero dei dati raccolti casualmente, più si incorre in una situazione di “normalità”, con il profilo della classica curva a campana (curva di Gauss). Se, per esempio, misurassimo l’altezza di un campione selezionato a caso all’interno di una certa popolazione, avremmo poche persone bassissime, poche altissime, e la maggior parte di media statura.

¹¹⁹ Nel caso di un numero dispari di valori si tratta di estrarre il valore che si colloca a metà (es. nella sequenza 3, 6, 66, 100, 104; la mediana è 66), nel caso di un numero pari si tratta di calcolare la media aritmetica tra i valori che si collocano a metà [es. nella sequenza 2, 3, 6, 66, 100, 104, la *mediana* è data da $(6+66)/2$, cioè $72/2$, ovvero 36].

¹²⁰ No caso de um número ímpar de valores, é necessário extrair o valor que está no meio da sequência (ex. na sequência 3, 6, 66, 100, 104; a mediana é 66), no caso de um número par, é necessário calcular a média aritmética entre os valores que estão no meio da sequência [ex. na sequência 2, 3, 6, 66, 100, 104, a *mediana* é calculada por $(6+66)/2$, ou seja, $72/2$, ou 36].

¹²¹ Nella sequenza 1, 4, 4, 5, 7, 7, 7, 8, la *moda* è 7.

¹²² Na sequência 1, 4, 4, 5, 7, 7, 7, 8, a *moda* é 7.

Una curva normale prevede 3 **deviazioni standard** su ciascun lato, ripartite simmetricamente dalla **media** (fig. 1). Il 68% dei dati rientra nella prima deviazione (± 1 deviazione standard); il 95% attorno alla seconda (± 2 deviazioni standard); il 99.7% attorno alla terza (± 3 deviazioni standard).

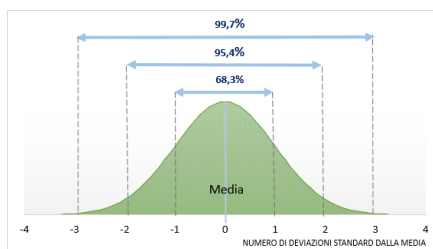


Fig. 1. *Curva normale e distribuzione*¹²³

Normalidade

É uma condição específica de **distribuição**, em que **média**, **mediana** e **moda** são equivalentes. Neste caso, temos uma simetria perfeita dos dados (com convergência para a parte central da curva). Quanto maior o número de dados coletados aleatoriamente, mais você se depara com uma situação de “normalidade”, com o contorno da curva clássica em forma de sino (curva de Gauss). Ao medir, por exemplo, a altura em uma amostra selecionada aleatoriamente em uma determinada população, teremos poucas pessoas

¹²³ L'immagine è stata tratta dal sito *Cube Investimenti* <<http://cubeinvestimenti.it/la-deviazione-standard-quello-devi-sapere/#jp-carousel-5245>>. Gentile concessione.

muito baixas, poucas muito altas, e a maior parte delas com estatura média.

Uma curva normal tem 3 **desvios padrão** de cada lado, divididos simetricamente pela **média** (fig. 1). 68% dos dados estão dentro do primeiro desvio (± 1 desvio padrão); 95% em torno do segundo (± 2 desvios padrão); 99.7% em torno do terceiro (± 3 desvios padrão).

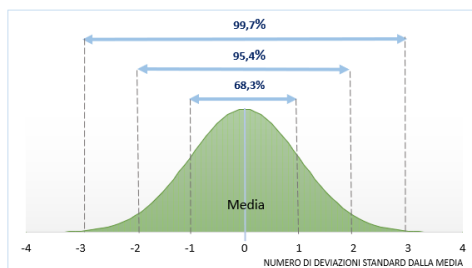


Fig. 1. *Curva normal e distribuição*¹²⁴

Punteggio zeta (z score)

Valore ottenuto mediante la divisione tra lo scarto del punteggio del candidato rispetto alla **media** e la **deviazione standard**. Può essere positivo o negativo a dipendere dal fatto che il punteggio di partenza sia superiore o inferiore alla media.

Mediante questa standardizzazione è possibile confrontare punteggi con **distribuzioni** differenti.¹²⁵

¹²⁴ A imagem foi obtida no site *Cube Investimenti* <<http://cubeinvestimenti.it/la-deviazione-standard-quello-devi-sapere/#jp-carousel-5245>>. Cortesia do site.

¹²⁵ Per approfondimenti, cfr. <<https://pellerrey.unisal.it/052006.pdf>>.

Escore-z (*z score*)

Valor obtido por divisão entre a diferença de pontuação de um candidato em relação à **média** geral dos candidatos e o **desvio padrão**. Pode ser positivo ou negativo dependendo se a pontuação inicial é maior ou menor do que a média geral dos candidatos.

Através dessa padronização é possível comparar pontuações com diferentes **distribuições**.¹²⁶

Significativtà statistica

Si tratta di un'indagine statistica inferenziale tesa ad accertare la possibilità di generalizzare (di far valere cioè oltre il campione di studio) la relazione tra due o più **variabili**.

L'indagine viene svolta mediante

- un *t test*, nel caso di due gruppi/variabili;¹²⁷

¹²⁶ Para mais informações, cfr. <<https://pellerer.unisal.it/052006.pdf>>.

¹²⁷ Si possono dare due tipi di *t test*: (a) *parametrici* e (b) *non parametrici*.

I *t test parametrici* sono adottati nel caso di una distribuzione **normale**, e si basano sulle differenze tra le **medie** dei due gruppi. A seconda dei casi possiamo avere:

- *test parametrico a campioni indipendenti*, nel caso si voglia comparare il comportamento di due gruppi in riferimento a una stessa esercitazione (per esempio: maschi e femmine alle prese con una stessa attività; oppure studenti che vivono nell'area urbana rispetto a quelli che vivono nell'area rurale, sempre in riferimento a una stessa attività);
- *test parametrico a campioni accoppiati*, nel caso si voglia comparare il comportamento di uno stesso gruppo alle prese con due attività differenti (per esempio, a uno stesso campione vengono assegnati compiti di lettura afferenti a due certificazioni diverse; oppure si analizzano i risultati raggiunti dallo stesso campione su due abilità differenti all'interno di una stessa prova: lettura e scrittura).

I *test non parametrici* sono adottati nel caso anche una sola delle due variabili presenti una **distribuzione** non normale e si basano sulla differenza di **mediana** (anziché di media). A seconda dei casi possiamo avere:

- *test non parametrico a campioni indipendenti* (Mann-Whitney U), come nel caso sopra descritto per i test parametrici;
- *test non parametrico a campioni correlati* (Wilcoxon), come nel caso sopra descritto relativo ai campioni accoppiati.

- un *test ANOVA*, nel caso di un numero maggiore di gruppi/variabili.¹²⁸

La significatività è definita per confutazione dell'*ipotesi nulla* (H_0), in virtù della quale *non* vi è differenza significativa tra due variabili.¹²⁹

Il valore di riferimento della significatività statistica è un *indice di probabilità* (p ; compreso tra **0-1**). Il valore soglia nelle scienze sociali è stabilito convenzionalmente allo **0.05**; in sostanza,

- con **$p > 0.05$** l'*ipotesi nulla* è confermata: non abbiamo l'evidenza di una differenza statisticamente significativa tra le due (o più)

¹²⁸ Il *test ANOVA* prevede il confronto tra una variabile dipendente (es. punteggio raggiunto da una certa popolazione in una certa esercitazione) e una variabile indipendente in più di due gruppi (es. la popolazione dei candidati divisi per fasce di età, o per regioni di provenienza, ecc.). Il *test ANOVA* ci informa se esiste una **differenza statistica significativa** (non imputabile al caso) nel comportamento dei diversi sottogruppi. Il *test ANOVA* presume che la **varianza** tra i gruppi sia omogenea (in caso contrario, occorre attuare procedure di calcolo alternative: *Brown-Forsythe* e *Welch*). Una volta accertata la **significatività** della differenza, un'indagine più dettagliata, in grado di localizzare dove esattamente si colloca la differenza tra i gruppi, è il *test post hoc* (la procedura di calcolo raccomandata, nel caso di omogeneità di **varianza**, è la *Scheffe* o *Tukey*; nel caso di disomogeneità di **varianza**, è la *Tamhane T2* o *Dunnnett T3*).

¹²⁹ Un procedimento per accertare la **significatività statistica**, applicabile a prescindere dal numero delle variabili, è pure il test del *Chi Quadro* (*Chi square*; vi si ricorre spesso nello studio del **comportamento differenziato degli item**). Il valore viene elaborato mediante calcoli riferiti alle differenze tra i dati ipotizzabili nel caso le variabili presentassero una stessa distribuzione e i dati osservati (<www.quadernodiepidemiologia.it/epi/assoc/chi_qua.htm>). Tabelle apposite permettono di calcolare il valore del *Chi Quadro* in funzione dei *gradi di libertà* <www.di-mgt.com.au/chisquare-table.html>.

I *gradi di libertà* sono definiti in Green con queste parole (2013: 92; il corsivo è nostro): “Il numero di valori che variano in relazione ad altri valori noti. Se, per esempio, noi abbiamo tre valori: $A + B = C$, una persona potrà attribuire qualsiasi valore ad A e a B , e il risultato sarà sempre C . Così, $4 + 3$ significherà che $C=7$; $8 + 5$ vorrà dire che $C=13$. Ma se C è un valore dato, mettiamo 100, allora solo uno degli altri valori (sia esso A o B) può variare, perché nel momento stesso in cui viene attribuito un valore A o a B , il terzo risulta automaticamente dalla sottrazione di C con il valore da noi attribuito. Per esempio: se $A = 35$ e $C = 100$, allora $B=65$. [Pensiamo a un altro caso:] $A + B + C = 100$; in questa circostanza, due valori ($A + B$, o $B + C$, o $A + C$) possono variare: abbiamo cioè due gradi di libertà. Posto invece che $A + B + C + D = 100$, sono tre i valori che possono variare, il che significa tre gradi di libertà, e così via”. In sostanza, i gradi di libertà si possono stabilire mediante la somma delle occorrenze meno un'unità ($n-1$).

variabili, ovvero generalizzabile oltre il campione di riferimento;

- con $p < 0.05$ l'*ipotesi nulla* è confutata: si può affermare che la differenza tra le due (o più) variabili è statisticamente significativa.

Significância estatística

Trata-se de um levantamento estatístico inferencial que permite verificar a possibilidade de generalizar (isto é, tirar conclusões para além da amostra do estudo) a relação entre duas ou mais **variáveis**.

O levantamento é realizado através de

- um *test t*, no caso de dois grupos/variáveis;¹³⁰
- uma *análise de variância* (ANOVA), para mais grupos/variáveis.¹³¹

¹³⁰ Dois tipos de *testes t* podem ser dados: (a) *paramétrico* e (b) *não paramétrico*.

Os *testes t paramétricos* são utilizados no caso de uma distribuição **normal** e têm por base as diferenças entre as **médias** dos dois grupos. Dependendo do caso, podemos ter:

- *teste paramétrico com amostras independentes*, se você quiser comparar o comportamento de dois grupos em relação a uma mesma atividade (por exemplo: homens e mulheres empenhados na mesma atividade; ou estudantes que vivem na área urbana em comparação com aqueles que vivem na área rural, sempre se referindo à mesma atividade);
- *teste paramétrico com amostras pareadas*, se você quiser comparar o comportamento do mesmo grupo empenhado em duas atividades diferentes (por exemplo, atribuímos à mesma amostra tarefas de leitura relacionadas a dois exames de proficiência diferentes; ou analisamos os resultados obtidos pela mesma amostra em duas habilidades diferentes dentro da mesma prova: leitura e escrita).

Os *testes não paramétricos* são utilizados no caso em que apenas uma das duas variáveis apresente uma **distribuição** não normal e têm por base a diferença de **mediana** (ao invés da média). Dependendo do caso, podemos ter:

- *testes não paramétricos com amostras independentes* (Mann-Whitney U), como descrito acima para testes paramétricos;
- *testes não paramétricos com amostras dependentes entre si* (Wilcoxon), como no caso acima descrito para amostras pareadas.

¹³¹ A *análise de variância ANOVA* baseia-se na comparação de uma variável dependente (ex. a nota obtida por uma determinada população em uma determinada atividade) com uma variável independente em mais de dois grupos (ex. a população de candidatos separados por três ou mais faixas etárias, ou por três ou mais regiões de origem, etc.). A *análise de variância ANOVA* informa se há diferença estatística **significativa** (não atribuível ao acaso) no comportamento

A significância é definida pela rejeição da *hipótese nula* (H_0), em virtude da qual *não* há diferença significativa entre duas variáveis.¹³²

O valor de referência da significância estatística é um *índice de probabilidade* (p ; entre **0-1**). O valor limite nas ciências sociais é convencionalmente fixado em **0.05**; basicamente,

- com $p > 0.05$ a *hipótese nula* está confirmada: não temos evidência de uma diferença estatisticamente significativa entre as duas (ou mais) variáveis, ou seja, não é possível generalizar para além da amostra de estudo;

dos diferentes subgrupos. A *análise de variância ANOVA* pressupõe que a **variância** entre os grupos seja homogênea (caso contrário, é necessário aplicar métodos de cálculo alternativos: *Brown-Forsythe* e *Welch*). Uma vez que tenha sido estabelecida a **significância** da diferença, o *teste post hoc* oferece uma investigação mais detalhada, capaz de localizar onde exatamente se situa a diferença entre os grupos (o procedimento de cálculo recomendado, no caso da homogeneidade da **variância**, é o *teste de Scheffé* ou de *Tukey*; no caso da falta de homogeneidade de **variância**, é o *teste de Tamhane T2* ou de *Dunnett T3*).

¹³² Um procedimento para verificar a **significância estatística**, aplicável independentemente do número de variáveis, é também o teste do *Qui-quadrado* (*Chi square*; é frequentemente utilizado no estudo do **comportamento diferenciado dos itens**). O valor é estabelecido através de cálculo das diferenças entre dados hipotéticos, se as variáveis tiverem a mesma distribuição, e dados observados (<www.quadernodiepidemiologia.it/epi/assoc/chi_qua.htm>). Tabelas específicas permitem calcular o valor do *Qui-quadrado* em função dos *graus de liberdade* <www.dimgt.com.au/chisquare-table.html>.

Os *graus de liberdade* são definidos por Green com estas palavras (2013: 92; o destaque em itálico é nosso): “O número de valores que varia em relação a outros valores conhecidos. Se, por exemplo, tivermos três valores: $A + B = C$, uma pessoa pode atribuir qualquer valor a A e B , e o resultado será sempre C . Assim, $4 + 3$ significará que $C=7$; $8 + 5$ significará que $C=13$. Mas se C tem um valor preestabelecido, digamos 100, então apenas um dos outros valores (seja A ou B) pode variar, porque ao mesmo tempo que um valor é atribuído a A ou B , o terceiro automaticamente resulta da subtração de C com o valor atribuído por nós. Por exemplo: se $A= 35$ e $C= 100$, então $B=65$. [Pensemos em outro caso:] $A + B + C = 100$; nesta circunstância, dois valores ($A + B$, ou $B + C$, ou $A + C$) podem variar: temos dois graus de liberdade. Outro caso ainda: $A + B + C + D = 100$, existem três valores que podem variar, o que significa três graus de liberdade, e assim por diante”. Basicamente, os graus de liberdade podem ser estabelecidos pela soma das ocorrências menos uma unidade ($n-1$).

- com $p < 0.05$ a hipótese nula é rejeitada: pode-se dizer que a diferença entre as duas (ou mais) variáveis é estatisticamente significativa.

Stima dell'effetto (*effect size*)

Il valore (calcolato diversamente a seconda che si abbia a che fare con test parametrici [con distribuzione **normale**] o non parametrici [con **curtosi** e **asimmetria** elevate]) ci informa della forza dell'associazione tra due variabili.

Il calcolo viene eseguito una volta appurata la **significatività statistica** e vale a indicare se la differenza tra le **variabili** è appunto rilevante (in effetti possiamo pur avere una differenza statisticamente significativa, ma di scarso rilievo). I valori dell'effetto sono riportati qui sotto.

ridotto	médio	ampio	Fonte
0.10	0.30	0.50	J. COHEN 1988 ²

Tamanho do efeito (*effect size*)

O valor (calculado diversamente dependendo se temos que lidar com testes paramétricos [com **distribuição normal**] ou não paramétricos [com alta **curtose** e **assimetria**]) indica a força de relação entre duas variáveis.

O cálculo é realizado tendo sido determinada a **significância estatística** e indica se a diferença entre as

variáveis é relevante (na verdade, podemos ter uma diferença estatisticamente significativa, mas de pouca importância). Os valores do tamanho do efeito são indicados abaixo.

Reduzido	médio	alto	Fonte
0.10	0.30	0.50	J. COHEN 1988 ²

Variable

Fattore oggetto di indagine.

In base al tipo di dati a cui tale fattore è riferito, possiamo distinguere:

- *variabile continua*, definita da dati che possono essere disposti all'interno di una scala lineare, con intervalli quantitativamente definibili;
- *variabile ordinale*, definita da dati che possono essere disposti secondo un certo ordine, senza la possibilità però di definire esattamente la differenza tra gli intervalli (es. Scala Likert);
- *variabile nominale*, definita dagli elementi che appartengono a una categoria (es. nazionalità, genere, età, ecc.).

In base alla relazione con altre variabili all'interno di un determinato studio, possiamo distinguere:

- *variabile dipendente*, la quale varia in rapporto a un'altra variabile;

- *variabile indipendente*, la quale si suppone non vari in rapporto all'altra variabile.¹³³

In base ai valori che la variabile può assumere, distinguiamo

- *variabile dicotomica* (con due valori assegnabili: 0, 1);
- *variabile politomica* (può assumere tre o più valori).

Un'analisi statistica può riguardare due variabili (es.: abilità del candidato/ difficoltà degli *item*) o più (es.: livello di competenza del candidato/ grado di difficoltà del compito di produzione scritta o orale/ grado di severità del correttore).

Variável

Fator a ser investigado.

Com base no tipo de dados a que este fator se refere, podemos distinguir:

- *variável contínua*, definida por dados que podem ser dispostos dentro de uma escala linear, com intervalos quantitativamente definíveis;
- *variável ordinal*, definida por dados que podem ser organizados em uma determinada ordem, mas sem a possibilidade de definir exatamente a diferença entre os intervalos (ex. Escala Likert);
- *variável nominal*, definida pelos elementos pertencentes a uma categoria (ex. nacionalidade, sexo, idade, etc.).

Com base na relação com outras variáveis dentro de um determinado estudo, podemos distinguir:

¹³³ Si pensi, per esempio, al punteggio raggiunto nella comprensione di un testo (*variabile dipendente*), rapportato allo **stile di apprendimento** dei candidati (*variabile indipendente*).

- *variável dependente*, que varia em relação à outra variável;
- *variável independente*, que se presume não variar em relação à outra variável.¹³⁴

Com base nos valores que a variável pode assumir, distinguimos

- *variável dicotômica* (pode assumir dois valores: 0, 1);
- *variável politômica* (pode assumir três ou mais valores).

Uma análise estatística pode abranger duas variáveis (ex.: habilidade do candidato/ dificuldade dos itens) ou mais (ex.: nível de competência do candidato/ grau de dificuldade da tarefa de produção escrita ou oral/ grau de severidade do corretor).

Varianza

È un *indice di distribuzione*, ottenuto mediante la divisione della somma del quadrato delle distanze dalla **media** dei singoli punteggi con il numero delle occorrenze meno un'unità [$\Sigma d^2 / (n-1)$].

Variância

É um **índice de distribuição**, que é obtido dividindo a soma do quadrado das distâncias da **média** de cada nota pelo número de ocorrências menos uma unidade [$\Sigma d^2 / (n-1)$].

¹³⁴ Pense, por exemplo, na nota alcançada na compreensão de um texto (*variável dependente*), relacionada ao **estilo de aprendizagem** dos candidatos (*variável independente*).

H. METAVALUTAZIONE

H. META-AVALIAÇÃO

Riassumiamo in questa sezione una serie di questioni filosofiche, psicologiche, pedagogiche e sociologiche inerenti l'atto del valutare.

Nesta seção, resumimos uma série de questões filosóficas, psicológicas, pedagógicas e sociológicas relativas ao ato de avaliar.

Affaticamento (cognitive fatigue)

Difficoltà di concentrazione legata alla lunghezza eccessiva della prova (cfr. ACKERMAN, KANFER, 2009).

Fadiga (cognitive fatigue)

Dificuldade de concentração devido à extensão da prova (cfr. ACKERMAN, KANFER, 2009).

Ansia da esame (test anxiety)¹³⁵

Stato di tensione che un'esercitazione/prova può procurare, connesso alla paura che la propria immagine sociale venga intaccata e/o al timore delle conseguenze legate al fallimento (DENSCOMBE, 2000).¹³⁶

¹³⁵ Detta anche *examinophobia* o *exam fever*.

¹³⁶ Attività di familiarizzazione con il *format* (**simulazioni**), ricognizioni metacognitive sul tipo di **attribuzioni causali** e, più in generale, ogni iniziativa che corrobori l'autostima possono aiutare il candidato a gestire le proprie emozioni (DAVIS, RINVOLUCRI, 2006).

*Estresse pré-prova (test anxiety)*¹³⁷

Estado de tensão que uma atividade/prova pode gerar, ligado ao medo de que a própria imagem social seja afetada e/ou ao medo das consequências do fracasso (DENSCOMBE, 2000).¹³⁸

Atteggiamento

Risposta emotiva generalizzata, sostenuta da **convinzioni**.

Un atteggiamento positivo coglie una prova come un'occasione per esibire quanto appreso e/o come una sfida personale; un atteggiamento negativo, al contrario, coglie la prova come una fonte di stress e/o una fastidiosa *routine*.

Atitude perante a prova

Resposta emotiva generalizada, apoiada por **convicções**.

Uma atitude positiva aproveita a prova como uma oportunidade de mostrar o que foi aprendido e/ou como um desafio pessoal; uma atitude negativa, ao contrário, compreende a prova como uma fonte de estresse e/ou uma rotina incômoda.

¹³⁷ Também chamado de *examinophobia* ou *exam fever*.

¹³⁸ Atividades de familiarização com o tipo de prova (**simulados**), atividades de reflexão metacognitiva sobre o tipo de **atribuição causal** e, de forma mais geral, qualquer iniciativa que fortaleça a autoestima, tudo isso pode ajudar o candidato a gerenciar suas emoções (DAVIS, RINVOLUCRI, 2006).

Atribuzione causale

Si tratta di una particolare **convinzione**, in virtù della quale l'individuo attribuisce l'origine dei suoi successi o insuccessi a un certo fattore. Le attribuzioni si distinguono in

- *interne*, se l'origine del successo/insuccesso coincide con il soggetto stesso (colto così come responsabile del proprio successo);
- *esterne*, se l'origine coincide con fonti esterne (per esempio, la severità del docente o la fortuna/sfortuna).

A fronte dell'insuccesso, attribuzioni esterne tendono a generare un **atteggiamento** fatalistico, mentre quelle interne tendono a stimolare un **atteggiamento** proattivo (la persona conta di potercela fare, una volta messi il dovuto impegno; cfr. BANDURA, 1977; 1997; WEINER, 1979).

Atribuição de causalidade

Trata-se de uma **convicção** específica, em virtude da qual o indivíduo atribui a origem de seus sucessos ou fracassos a um determinado fator. As atribuições são divididas em

- *internas*, se a origem do sucesso/fracasso coincidir com o próprio sujeito (que é responsável pelo seu próprio sucesso);
- *externas*, se a origem coincide com fontes externas (por exemplo, a severidade do docente ou sorte/azar).

Diante do fracasso, atribuições externas tendem a gerar uma **atitude fatalística perante a prova**, enquanto as internas tendem a estimular uma **atitude proativa perante a prova** (a pessoa acredita que consegue fazer, depois de dedicar o devido empenho; cfr. BANDURA, 1977; 1997; WEINER, 1979).

Autoregolazione

Capacità di gestire le condizioni che consentono di ottimizzare il proprio apprendimento (cfr. BOEKAERTS *et al.*, 2000; KUHL, 2000). Un apprendente autoregolato riflette sui propri processi cognitivi (cfr. **strategie metacognitive**) e sulle proprie motivazioni (cfr. **strategie regolative**).

Autorregulação

Capacidade de lidar com as condições que permitem otimizar a própria aprendizagem (cfr. BOEKAERTS *et al.*, 2000; KUHL, 2000). Um aluno autorregulado reflete sobre seus próprios processos cognitivos (cfr. **estratégias metacognitivas**) e motivações (cfr. **estratégias reguladoras**).

Controvalutazione

Atteggiamento negativo nei confronti della valutazione, presente in alcuni contesti educativi. La valutazione è colta come un'azione che può inibire la formazione di una **motivazione** interna.

Antiavaliação

Atitude negativa **perante a prova**, presente em alguns contextos educacionais. A avaliação é entendida como uma ação que pode inibir a formação de uma **motivação** interna.

Convinzione

Idea preconstituita su cos'è la valutazione, come si deve valutare, qual è il ruolo del valutatore e qual è quello del *test taker*.

Le convinzioni sono relativamente stabili;¹³⁹ inducono ad avere determinati **atteggiamenti**; possono emergere grazie a ricognizioni metacognitive.¹⁴⁰

Convicção

Ideia pré-estabelecida do que é avaliação, de como se deve avaliar, de qual é o papel do avaliador e de qual é o do examinando.

As convicções são relativamente estáveis;¹⁴¹ induzem determinadas **atitudes perante a prova**; podem ser investigadas através de ferramentas metacognitivas.¹⁴²

¹³⁹ Scrivono Mariani e Pozzo (2002: 49): “Pur essendo altamente soggettive, e quindi non necessariamente corrispondenti alla «realtà delle cose» o a «teorie» accreditate, le convinzioni hanno per l'individuo [...] la forza di una «filosofia personale» molto strutturante e motivante. In altre parole, la mia personale, soggettiva percezione della realtà costituisce per me «la realtà» *tout court*. [...] Dato il loro carattere di forze strutturanti la propria «visione del mondo», le convinzioni sono relativamente stabili e non facilmente o rapidamente passibili di modificazioni. [...] Un modello interpretativo costruito magari in anni di esperienze non può essere modificato se non con nuove, forti, ripetute esperienze di segno diverso”.

¹⁴⁰ Mariani e Pozzo commentano (2002: 51-53): “Solo la condivisione esplicita di contenuti cognitivi e affettivi inesplorati permette di riconoscere questi contenuti come appartenenti alla propria esperienza, di confrontarli con una realtà «esterna» fatta di contenuti altri, e, se ritenuto necessario o opportuno, di entrare nell'ottica di modificare o aggiornare le proprie posizioni. Questionari, interviste, scale di valutazione, diari sono solo alcuni esempi di strumenti utilizzabili a questo scopo”.

¹⁴¹ Mariani e Pozzo escrevem (2002: 49): “Embora sejam altamente subjetivas, e, portanto, não necessariamente correspondentes à «realidade das coisas» ou a «teorias» consolidadas, as convicções têm para o indivíduo [...] a força de uma «filosofia pessoal» muito estruturante e motivadora. Em outras palavras, a minha percepção pessoal e subjetiva da realidade constitui para mim «a realidade» *tout court*. [...] Considerando a sua natureza como força estruturante da própria «visão de mundo», as convicções são relativamente estáveis e não estão sujeitas facilmente ou rapidamente a mudanças. [...] Um modelo interpretativo construído talvez em anos de experiência não pode ser modificado a não ser com experiências novas, fortes e repetidas de outro caráter”.

¹⁴² Mariani e Pozzo comentam (2002: 51-53): “Só a partilha explícita de conteúdos cognitivos e afetivos inesplorados permite conhecer estes conteúdos como pertencentes à própria experiência, compará-los com a realidade «externa» feita de outros conteúdos, e, se necessário ou apropriado, assumir a perspectiva de modificar ou atualizar os próprios posicionamentos.

Effectance

È un tipo di **motivazione** intrinseca, legato al piacere/ la volontà di esercitare un controllo sull'ambiente (WHITE, 1959; HARTER, 1978). È frutto dell'interiorizzazione di un'approvazione esterna; si consolida cioè per via di *feedback* positivi/ costruttivi ripetuti.¹⁴³ Sostiene un **atteggiamento** positivo nei confronti di una prova.

Effectance

É um tipo de **motivação** intrínseca, ligada ao prazer/ à vontade de exercer o controle sobre a situação (WHITE, 1959; HARTER, 1978). É o resultado da internalização de uma aprovação externa; ou seja, consolida-se através de repetidos *feedbacks* positivos/ construtivos.¹⁴⁴ Favorece uma **atitude** positiva **perante a prova**.

Questionários, entrevistas, escalas de avaliação, diários são apenas alguns exemplos de ferramentas que podem ser usadas para este fim”.

¹⁴³ Scrivono De Beni e Moè (2000: 56-57; il corsivo è nostro): “Se il bambino viene sostenuto [... nei] tentativi di padronanza e ottiene rinforzi positivi o approvazione [...], tende a sviluppare un sistema di autoricompenza. *Tale sistema di autoricompenza rende sempre meno importante l'approvazione esterna dell'adulto e consente di sviluppare obiettivi di padronanza [...].* Se il bambino, invece, viene scoraggiato o disapprovato nei tentativi di padronanza, [...] tende a sviluppare un bisogno di approvazione esterna che lo porta a sentirsi dipendente dall'adulto e a porsi obiettivi di approvazione e di prestazione caratterizzati dal desiderio di mostrare le proprie abilità e dal complementare timore di sentirsi incapace. Tutto questo fa sentire il bambino meno competente e più controllato dall'esterno e comporta una diminuzione della motivazione della *effectance*. La bassa percezione di competenza può, inoltre, generare il timore di dimostrarsi incapace. Questo timore, a sua volta, produce esperienze emotive negative, quali l'ansia, che portano a una ulteriore diminuzione di *effectance*, innestando un processo circolare per cui *i fallimenti generano ansia di fronte alle situazioni di apprendimento future e fanno vivere la sensazione di non essere competenti?*”.

¹⁴⁴ De Beni e Moè escrivem (2000: 56-57; o itálico é nosso): “Se a criança é apoiada [... em] tentativas de domínio e recebe reforço positivo ou aprovação [...], ela tende a desenvolver um sistema de autoricompenza. *Esse sistema de autoricompenza torna a aprovação externa do adulto cada vez menos importante e permite desenvolver metas de domínio [...].* Se a criança, por outro lado, é desencorajada ou reprovada em tentativas de domínio, [...] tende a desenvolver uma necessidade de aprovação externa que a leva a se sentir dependente do adulto e a estabelecer metas de aprovação e desempenho caracterizados pelo desejo de mostrar suas habilidades e pelo medo concomitante de se sentir incapaz. Tudo isso faz com que a criança se sinta menos competente e mais regulada pelo exterior e resulta em uma diminuição da motivação de *effectance*. A baixa percepção de competência também pode dar origem ao medo de se mostrar incapaz. Esse medo, por sua vez, produz experiências emocionais negativas, como a ansiedade, que levam a uma diminuição ainda maior da *effectance*, desencadeando um processo circular

Etica (della valutazione)

Serie di attenzioni volte a garantire che la prova agisca nel rispetto dei bisogni dei candidati (cfr. SHOHAMY, 1997; HAMP-LYONS, 1998; BARNI 2010; FAN *et al.*, 2019).

Ética (da avaliação)

Série de medidas para garantir que a prova respeite as necessidades dos candidatos (cfr. SHOHAMY, 1997; HAMP-LYONS, 1998; BARNI 2010; FAN *et al.*, 2019).

Ipervalutazione

Mentre la **controvalutazione** rimanda a una filosofia della libertà (lo studente impara facendo, senza la necessità di un giudizio esterno), l'ipervalutazione rimanda a una filosofia del controllo (*hyper-accountability*; MANSELL, 2007), contraddistinta dall'uso costante di test volti ad accertare l'assimilazione di microsegmenti di conoscenza/ competenza. L'eccesso di verifica tuttavia corre il rischio di (TESSARO, 2004: 5) “trasforma[re] l'istruzione in mero addestramento”: la sperimentazione tende ad essere scoraggiata; il discente può provare una sensazione di oppressione.

Nell'ambito dell'educazione linguistica, contesti ipervalutativi si sposano con un **silabo** grammatico-lessicale, millimetricamente definito; l'esercizio delle abilità produttive viene ridotto ad attività di puro reimpiego di lessico e strutture.

em que as falhas geram ansiedade diante de futuras situações de aprendizagem e fazem com que as pessoas se sintam não competentes”.

Excesso de avaliações Enquanto a **antiavaliação** remete a uma filosofia da liberdade (o estudante aprende fazendo, sem a necessidade de uma avaliação externa), o excesso de avaliações remete a uma filosofia de controle (*hyper-accountability*; MANSELL, 2007), com a utilização constante de testes para determinar a assimilação de microsegmentos de conhecimento/competência. Esse excesso, no entanto, corre o risco de (TESSARO, 2004: 5) “transforma[r] a educação em mero treinamento”: a experimentação tende a ser desencorajada; o discente pode experimentar um sentimento de opressão. No campo da educação linguística, contextos com excesso de avaliações combinam com **componentes curriculares** léxico-gramaticais, definidas milimetricamente; o exercício de habilidades produtivas é reduzido a atividades de pura reutilização de léxico e estruturas.

Motivazione

Dedizione, interesse e persistenza nello studio. Si tratta di un **costrutto** dinamico; è sensibile, in particolare, all’interazione tra individuo e contesto.¹⁴⁵

Nella psicologia sussistono varie distinzioni in merito alla motivazione. I tipi principali di motivazione sono:

- la *motivazione estrinseca* (legata a stimoli esterni: ottenere vantaggi o evitare conseguenze negative);
- la *motivazione intrinseca* (legata a stimoli interni; cfr. motivazione di **effectance**).¹⁴⁶

¹⁴⁵ Per approfondimenti circa il ruolo della motivazione nell’apprendimento di una lingua straniera, raccomandiamo la lettura di Gardner, 1985; Bernaus, Wilson, 2009; Dörnyei, Ushioda, 2021³.

¹⁴⁶ La distinzione tra i due non è netta; nel corso del tempo la motivazione può mutare di segno.

Motivação

Dedicação, interesse e persistência no estudo. Trata-se de um **constructo** dinâmico; é sensível, em particular, à interação entre indivíduo e contexto.¹⁴⁷

Em psicologia existem várias distinções sobre motivação. Os principais tipos de motivação são:

- a *motivação extrínseca* (associada a estímulos externos: obter vantagens ou evitar consequências negativas);
- a *motivação intrínseca* (associada a estímulos internos; cfr. motivação de **effectance**).¹⁴⁸

Stile di apprendimento

Preferenza accordata dall'apprendente a determinate attività/ strategie (CADAMURO, 2004).

Per esempio, alcuni apprendenti possono prediligere attività strutturate, mentre altri sono maggiormente disposti verso attività divergenti. È un dato di cui si può tener conto nel contesto di una **valutazione personalizzata**.¹⁴⁹

Estilo de aprendizagem

Preferência por certas atividades/ estratégias por parte do aluno (CADAMURO, 2004).

¹⁴⁷ Para mais informações sobre o papel da motivação na aprendizagem de uma língua estrangeira, recomendamos a leitura de Gardner, 1985; Bernaus, Wilson, 2009; Dörnyei, Ushioda, 2021³.

¹⁴⁸ A distinção entre as duas não é clara; com o tempo a motivação pode mudar de aspecto.

¹⁴⁹ Alcuni studiosi sono convinti, ad ogni modo, che non si tratti di una caratteristica stabile. La persona può dimostrare, per esempio, un comportamento più olistico (con una preferenza cioè verso l'insieme dei dati) o analitico (con l'attenzione verso i particolari) in funzione dell'interesse e della competenza. In altre parole, gli studenti si dimostrerebbero più riflessivi e attenti ai dettagli mentre lavorano nelle discipline che li interessano di più (cfr. GAZZANIGA, 1985; GARDNER *et al.*,1998; in merito alla flessibilità degli stili, cfr. anche LEAMNISON, 1999).

Por exemplo, alguns alunos podem preferir atividades estruturadas, enquanto outros estão mais dispostos a realizar atividades divergentes. É um fato que pode ser levado em conta no contexto de uma **avaliação personalizada**.¹⁵⁰

Valori Idee e principi che regolano la redazione di una prova linguistica. Ogni forma di valutazione esprime, in effetti, una visione di mondo e di società (FULCHER, 2015).

Valores Ideias e princípios que regem a elaboração de uma prova de língua. Cada forma de avaliação expressa, de fato, uma visão de mundo e da sociedade (FULCHER, 2015).

Valutazione dei valutatori Processo mediante il quale l'attività di un agente di valutazione (dal singolo insegnante al centro certificatore) è monitorata da un soggetto terzo.

Avaliação dos avaliadores Procedimento através do qual se monitora a atividade de um agente de avaliação (seja um professor, seja um ente certificador).

¹⁵⁰ Alguns estudiosos estão convencidos, no entanto, que esta não é uma característica estável. A pessoa pode demonstrar, por exemplo, um comportamento mais holístico (isto é, com uma preferência em relação ao conjunto de dados) ou analítico (com atenção aos detalhes) dependendo do interesse e da competência. Em outras palavras, os estudantes estariam mais conscientes e atentos aos detalhes trabalhando nas disciplinas que mais lhes interessam (cfr. GAZZANIGA, 1985; GARDNER *et al.*, 1998; sobre a flexibilidade dos estilos, cfr. também LEAMNISON, 1999).

Parte III
Documenti e Risorse

Parte III
Documentos e Recursos

DOCUMENTI EUROPEI

DOCUMENTOS EUROPEUS

Illustriamo alcuni documenti elaborati dal Consiglio d'Europa per l'insegnamento delle lingue e la valutazione linguistica.

Apresentamos alguns documentos elaborados pelo Conselho Europeu para o ensino de línguas e avaliação linguística.

Companion (2020) Aggiornamento del **Quadro di Riferimento**, con l'approfondimento di alcuni descrittori (in particolari quelli legati alla fonologia). Il documento dedica una particolare attenzione, inoltre, alla competenza plurilingue, alle abilità di mediazione, all'abilità di lettura di testi letterari e alla definizione del livello pre-A1.

Companion (2020) Atualização do **Quadro Comum Europeu**, com o aprofundamento de alguns descritores (em particular os relacionados à fonologia). O documento também dedica atenção especial à competência plurilíngue, às habilidades de mediação, à habilidade de leitura de textos literários e à definição do nível pré-A1.

Portfolio Europeo delle Lingue (2000)

Documento che orienta la progettazione di una raccolta delle prestazioni dell'allievo (chiamata appunto

portfolio). La funzione di tale raccolta è quella di favorire l'autovalutazione dell'allievo (cfr. LITTLE, PERCLOVA, 2001).

Portfólio Europeu de Línguas (2000)

Documento que ajuda a realizar uma coletânea do desempenho do aluno (chamada, justamente, de **portfólio**). A função desta coletânea é incentivar a autoavaliação do aluno (cfr. LITTLE, PERCLOVA, 2001).

Profili di varie lingue (dal 2005 in poi)

Documenti riferiti a varie lingue europee; declinano i descrittori del **Quadro** in contenuti linguistici, dando risposta a domande del tipo: *quali testi, quali funzioni, quali elementi morfosintattici, quale lessico l'allievo è in grado di produrre a un certo livello?*

Per un prospetto generale dei documenti, si acceda al seguente *link*: <www.coe.int/en/web/common-european-framework-reference-languages/reference-level-descriptions-rlds-developed-so-far>

Referenciais de várias línguas (de 2005 em diante)

Documentos referentes a várias línguas europeias; expressam os descritores do **Quadro** em conteúdos linguísticos, dando respostas a perguntas como: *que textos?, que funções?, que elementos morfosintáticos?, que léxico o aluno é capaz de produzir em um determinado nível?*

Para uma visão geral dos documentos, visite o seguinte *link*: <www.coe.int/en/web/common-european-

Quadro Comune Europeo di Riferimento per le Lingue (2001)

Documento volto a orientare insegnanti e valutatori mediante la definizione di sei livelli di competenza linguistico-comunicativa, secondo coordinate strategico-testuali:

- livello di contatto (*Breakthrough* – A1);
- livello di sopravvivenza (*Waystage* – A2);
- livello soglia (*Threshold* – B1);
- livello progresso (*Vantage* – B2);
- livello dell'efficacia (*Proficiency* – C1);
- livello della padronanza (*Mastery* – C2)

Quadro Europeu Comum de Referência para as Línguas (2001)

Documento destinado a orientar professores e avaliadores através da definição de seis níveis de competência linguístico-comunicativa, de acordo com coordenadas estratégico-textuais:

- nível de descoberta (*Breakthrough* – A1);
- nível plataforma (*Waystage* – A2);
- nível limiar (*Threshold* – B1);
- nível independente (*Vantage* – B2);
- nível de proficiência (*Proficiency* – C1);
- nível de domínio pleno (*Mastery* – C2)

RISORSE

RECURSOS

Presentiamo articoli, interventi video, e strumenti in generale che possono contribuire alla formazione del valutatore.

Apresentamos artigos, vídeos e ferramentas em geral que podem contribuir para a formação do avaliador.

Risorse ALTE Documenti presenti nel sito dell'*Association of Language Testers in Europe* <www.alte.org/Resources-ALTE>

Recursos ALTE Documentos disponíveis no site da *Association of Language Testers in Europe* <www.alte.org/Resources-ALTE>

Risorse British Council

Video divulgativi sui concetti-chiave della valutazione linguistica <www.britishcouncil.org/exam/aptis/research/assessment-literacy>.

Documenti elettronici scaricabili
<<https://www.britishcouncil.org/exam/aptis/research/publications>>.

Recursos British Council

Vídeos informativos sobre os principais conceitos de avaliação linguística <www.britishcouncil.org/exam/aptis/research/assessment-literacy>.

Documentos em versão digital para baixar
<<https://www.britishcouncil.org/exam/aptis/research/publications>>.

Risorse EALTA Documenti presenti sul sito dell'*European Association for Language Testing and Assessment*
<www.ealta.eu.org/resources.htm>.

Risorse EALTA Documentos disponíveis no site da *European Association for Language Testing and Assessment*
<www.ealta.eu.org/resources.htm>.

Risorse LATEFL TEASIG

Documenti presenti sul sito del gruppo *Testing, Assessment and Evaluation* della *International Association of Teachers of English as a Foreign Language*
<<https://tea.iatefl.org/>>.

Risorse LATEFL TEASIG

Documentos disponíveis no site do grupo *Testing, Assessment and Evaluation* da *International Association of Teachers of English as a Foreign Language*
<<https://tea.iatefl.org/>>.

Risorse ILTA

Video divulgativi e *podcast* di approfondimento messi a disposizione dall'*International Language Testing Association* (<<http://languagetesting.info>>). I 21 video, in particolare, valgono da glossario audiovisuale relativo alla valutazione linguistica:

- *Validity* (Bernard Spolsky)

- *Reliability* (Alan Davies)
- *Test Specifications* (Charles Alderson)
- *Item Writing* (Charles Stansfield)
- *Pretesting* (J. D. Brown)
- *Reading* (Caroline Clapham)
- *Writing* (Liz Hamp-Lyons)
- *Listening* (Gary Buck)
- *Speaking* (Glenn Fulcher)
- *Vocabulary* (John Read)
- *Integrated Assessment* (Lia Plakans)
- *LSP Testing* (Dan Douglas)
- *Test Taking Strategies* (Andrew Cohen)
- *Impact* (Elana Shohamy)
- *Statistics in Language Testing* (Fred Davidson)
- *Item Response Theory* (John de Jong)
- *Structural Equation Modeling* (Ardeshir Geranpayeh)
- *Technology in Language Testing* (Carol Chapelle)
- *Automated Scoring* (Xiaoming Xi)
- *Rater Bias* (Paula Winke and Susan Gass)
- *Assessing Pronunciation* (Talia Isaacs)

Recursos ILTA

Vídeos informativos e *podcasts* de aperfeiçoamento disponibilizados pela *International Language Testing Association* (<<http://languagetesting.info>>). Os 21 vídeos constituem um glossário audiovisual sobre a avaliação linguística:

- *Validity* (Bernard Spolsky)
- *Reliability* (Alan Davies)

- *Test Specifications* (Charles Alderson)
- *Item Writing* (Charles Stansfield)
- *Pretesting* (J. D. Brown)
- *Reading* (Caroline Clapham)
- *Writing* (Liz Hamp-Lyons)
- *Listening* (Gary Buck)
- *Speaking* (Glenn Fulcher)
- *Vocabulary* (John Read)
- *Integrated Assessment* (Lia Plakans)
- *LSP Testing* (Dan Douglas)
- *Test Taking Strategies* (Andrew Cohen)
- *Impact* (Elana Shohamy)
- *Statistics in Language Testing* (Fred Davidson)
- *Item Response Theory* (John de Jong)
- *Structural Equation Modeling* (Ardeshir Geranpayeh)
- *Technology in Language Testing* (Carol Chapelle)
- *Automated Scoring* (Xiaoming Xi)
- *Rater Bias* (Paula Winke and Susan Gass)
- *Assessing Pronunciation* (Talia Isaacs)

Risorse INTO EUROPE Project

4 volumi liberamente accessibili, aventi per oggetto la valutazione delle abilità linguistiche ed elaborati in seno alla riforma dell'esame di maturità in Ungheria, sotto la guida del British Council, in accordo con l'ente ministeriale ungherese per l'educazione OKI. I destinatari sono gli insegnanti di scuola secondaria; il taglio è divulgativo, la prosa è chiara.

<https://www.lancaster.ac.uk/fass/projects/examreform/Pages/IE_Writing.html>

Recursos INTO EUROPE Project

4 volumes de livre acesso sobre o tema da avaliação de habilidades linguísticas. Foram preparados como parte da reforma do exame de conclusão do ensino médio na Hungria, sob a orientação do British Council e de acordo com o Ministério da Educação húngaro OKI. Destinam-se aos professores do ensino médio; têm caráter divulgativo, o texto é claro. <https://www.lancaster.ac.uk/fass/projects/examreform/Pages/IE_Writing.html>

Risorse LAALTA

Raccolta di articoli dei membri della *Latin American Association for Language Testing and Assessment*
<<http://laalta.org/>>

Risorse LAALTA

Coletânea de artigos dos membros da *Latin American Association for Language Testing and Assessment*
<<http://laalta.org/>>

Risorse LTRG Lancaster

Raccolta di articoli, video e tesi elaborati in seno al *Language Testing Research Group* dell'Università di Lancaster, UK. <<http://wp.lancs.ac.uk/ltrg>>

Risorse LTRG Lancaster

Coletânea de artigos, vídeos e teses desenvolvidas no *Language Testing Research Group* da Universidade de Lancaster, UK. <<http://wp.lancs.ac.uk/ltrg>>

Riviste specializzate

Riviste dedicate al tema della valutazione linguistica:

- *Assessing Writing*
- *International Journal of Language Testing and Assessment*
- *Journal of Language Testing & Assessment*
- *Language Assessment Quarterly*
- *Language Testing*
- *Language Testing in Asia*
- *Melbourne Papers in Language Testing*

Altre riviste nelle quali vengono pubblicati articoli legati alla valutazione linguistica:

- *Assessment in Education: Principles, Policy & Practice*
- *Assessment and Evaluation in Higher Education*
- *Educational Evaluation and Policy Analysis*
- *Educational Measurement: Issues and Practice*
- *Educational and Psychological Measurement*
- *Evaluation and Research in Education*
- *Journal of Educational Measurement*
- *Practical Assessment, Research and Evaluation*
- *Trabalhos em Linguística Aplicada*

Revistas especializadas

Revistas dedicadas ao tema da avaliação linguística:

- *Assessing Writing*
- *International Journal of Language Testing and Assessment*
- *Journal of Language Testing & Assessment*
- *Language Assessment Quarterly*

- *Language Testing*
- *Language Testing in Asia*
- *Melbourne Papers in Language Testing*

Outras revistas em que artigos relacionados à avaliação linguística foram publicados:

- *Assessment in Education: Principles, Policy & Practice*
- *Assessment and Evaluation in Higher Education*
- *Educational Evaluation and Policy Analysis*
- *Educational Measurement: Issues and Practice*
- *Educational and Psychological Measurement*
- *Evaluation and Research in Education*
- *Journal of Educational Measurement*
- *Practical Assessment, Research and Evaluation*
- *Trabalhos em Linguística Aplicada*

Sito prof. G. Porcelli Varie pubblicazioni di libero accesso
<www.gporcelli.it/portent/LTeV.htm>

Site prof. G. Porcelli Várias publicações de livre acesso
<www.gporcelli.it/portent/LTeV.htm>

Parte IV
Appendici e riferimenti bibliografici

Parte IV
Apêndices e bibliografia

Appendice 1. Griglie analitiche per la valutazione della produzione linguistica

Illustriamo a seguire due griglie analitiche, una in italiano e una in portoghese (sono riportate nella loro lingua originale; non sono state soggette a traduzione). La prima, riportata qui sotto, è stata elaborata dalla *Secretaria Municipal de Educação* del Comune di Rio de Janeiro (I bimestre 2014), ed è finalizzata a valutare le competenze di scrittura di ragazzini del IX anno di studio (corrispondente alla III media del sistema scolastico italiano).

ASPECTOS	INDICADORES	PONTUAÇÃO
ESTRUTURA	<p>Apresentar elementos do texto, utilizando as características específicas da tipologia nos locais adequados e realizando a devida paragrafação.</p> <p>3 itens:</p> <ul style="list-style-type: none"> • Introdução; • Desenvolvimento; • Conclusão. 	<p>2,5 = Apresenta os 3 itens bem definidos, com uso adequado dos parágrafos.</p> <p>2,0 = Apresenta os 3 itens bem definidos, mas com falhas na paragrafação.</p> <p>1,5 = Apresenta os 3 itens, mas em um único parágrafo.</p> <p>1,0 = Apresenta 2 itens, mas com falhas na paragrafação.</p> <p>0,5 = Apresenta 1 ou dois itens, mas em um único parágrafo.</p> <p>0 = Não atende.</p>
COERÊNCIA	<p>Compreender a proposta de produção textual e desenvolver o tema, estabelecendo uma comunicação entre emissor e receptor por meio da elaboração de um texto estruturado.</p>	<p>2,5 = Narra a história, abordando os 4 itens, com coerência e clareza.</p> <p>2,0 = Narra a história, abordando os 4 itens, com algumas falhas na coerência ou na clareza.</p> <p>1,5 = Narra a história, abordando apenas 3 itens, porém escreve com coerência e clareza.</p>

	<p>4 itens:</p> <ul style="list-style-type: none"> • Apresentação; • Conflito gerador; • Clímax; • Desfecho. 	<p>1,0 = Narra a história, abordando apenas 3 itens, porém escreve com falhas na coerência ou na clareza.</p> <p>0,5 = Narra a história, abordando apenas 1 ou 2 itens, com pouca consistência, apresentando falhas na coerência e na clareza.</p> <p>0 = Não consegue formar um texto.</p>
COESÃO	<p>Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da produção textual (Recursos coesivos: palavras ou expressões utilizadas para ligar as partes do texto ou para evitar repetições desnecessárias).</p>	<p>2,5 = Articula <u>adequadamente</u> as partes do texto, sem apresentar problemas na utilização de recursos coesivos.</p> <p>2,0 = Articula <u>adequadamente</u> as partes do texto, podendo apresentar eventuais problemas na utilização de recursos coesivos.</p> <p>1,5 = Articula <u>razoavelmente</u> as partes do texto, mas apresenta problemas na utilização de recursos coesivos.</p> <p>1,0 = Articula <u>razoavelmente</u> as partes do texto, mas apresenta uso excessivo de repetições desnecessárias.</p> <p>0,5 = Articula <u>precarosamente</u> as partes do texto, devido a problemas frequentes na utilização de recursos coesivos.</p> <p>0 = <u>Não</u> articula as partes do texto.</p>
USO DA LÍNGUA	<p>Utilizar a escrita como sistema de representação, respeitando os seguintes aspectos do código linguístico:</p> <ul style="list-style-type: none"> • uso adequado do vocabulário; • ortografia – relação 	<p>2,5 = Demonstra <u>excelente</u> domínio do código linguístico, respeitando os aspectos ortográficos, de pontuação e de concordância.</p> <p>2,0 = Demonstra <u>bom</u> domínio do código linguístico, respeitando os aspectos ortográficos, de pontuação e de concordância.</p> <p>1,5 = Demonstra <u>bom</u> domínio do código</p>

	<p>grafema/fonema, espaçamento entre palavras, uso de maiúsculas e acentuação;</p> <ul style="list-style-type: none"> • pontuação – ponto final, ponto de exclamação, ponto de interrogação, vírgula; • concordância nominal e verbal (relações morfossintáticas). 	<p>linguístico, respeitando os aspectos de concordância, mas com algumas dificuldades em pontuação ou acentuação.</p> <p>1,0 = Demonstra <u>bom</u> domínio do código linguístico, respeitando os aspectos ortográficos ou de pontuação, mas com dificuldades na concordância.</p> <p>0,5 = Demonstra conhecimento <u>reduzido</u> do código linguístico, apresentando dificuldades com relação à pontuação, à acentuação e à concordância.</p> <p>0 = Não apresenta conhecimento mínimo do código linguístico.</p>
--	--	--

Tab. 1. Griglia analítica di valutazione della produzione scritta
(Secretaria Municipal da Educação, Prefeitura do Rio de Janeiro)¹⁵¹

La seconda è usata in sede di valutazione dell'interazione orale nell'esame di certificazione di italiano come lingua straniera PLIDA, al livello A2 (BARIVIERA *et al.*, 2021).

¹⁵¹ Per griglie di valutazione della produzione testuale elaborate dalla stessa istituzione, cfr. <<https://onedrive.live.com/view.aspx?resid=9A0409FEB0892781239&cid=09a0409feb089278&app=WordPdf>> (01.12.2021).

Para outras tabelas de avaliação da produção textual produzidas pela mesma instituição, cfr. <<https://onedrive.live.com/view.aspx?resid=9A0409FEB0892781239&cid=09a0409feb089278&app=WordPdf>>(01.12.2021).

	EFFICACIA COMUNICATIVA	INTERAZIONE	LESSICO	GRAMMATICA	PRONUNCIA
10	Realizza i compiti assegnati in modo pienamente soddisfacente. Produce sequenze più lunghe, utilizzando semplici meccanismi di coesione. Nonostante esitazioni e false partenze, il discorso può essere seguito con uno sforzo contenuto.	Risponde agli stimoli dell'interlocutore in maniera adeguata. È in grado di chiedere ripetizioni e chiarimenti in modo molto semplice. Cerca di usare qualche formula di cortesia.	Repertorio pienamente soddisfacente in rapporto alla consegna. Gli errori lessicali possono essere frequenti, ma il messaggio è generalmente chiaro. Le strategie di compensazione, se presenti, sono efficaci.	Padroneggia le strutture previste per il livello. Nonostante la presenza di errori, il messaggio è ben comprensibile.	L'interlocutore fa poca fatica a interpretare i passaggi in cui la pronuncia è scorretta.
9					
8	Realizza i compiti assegnati in modo adeguato anche se alcuni punti possono essere sviluppati meno di altri. Formule enunciate brevi e semplici, con qualche forma di coesione. Esitazioni e false partenze sono frequenti e a tratti rendono il discorso faticoso da seguire.	Risponde agli stimoli dell'interlocutore in maniera quasi sempre adeguata. È in grado di chiedere ripetizioni e chiarimenti in modo molto semplice. Applica in modo adeguato le principali formule di apertura e chiusura della conversazione.	Repertorio adeguato alla consegna. Gli errori lessicali possono rendere alcuni passaggi poco chiari. Le strategie di compensazione, se presenti, non sono sempre efficaci.	Padroneggia abbastanza bene le strutture previste per il livello. Errori frequenti possono rendere la comprensione del messaggio occasionalmente difficoltosa.	In alcuni passaggi l'interlocutore fa fatica a interpretare il messaggio a causa della pronuncia scorretta.
7					
6	Realizza i compiti assegnati in parte o in modo approssimativo. Si esprime con una serie di enunciati giustapposti; il legame logico tra gli enunciati è comunque chiaro.	L'interazione procede grazie agli sforzi dell'interlocutore. È in grado di segnalare che ha bisogno di ripetizioni e	Repertorio appena sufficiente in rapporto alla consegna. Numerosi errori lessicali possono rendere faticosa	Padroneggia sufficientemente le strutture previste per il livello. La comprensione del messaggio può essere compromessa in più punti.	L'interlocutore fa costantemente fatica a interpretare il messaggio a causa della pronuncia scorretta.

5	Esitazioni e false partenze sono frequenti e rendono il discorso faticoso da seguire.	chiarimenti in modo molto semplice anche non verbale. Applica con qualche incertezza le principali formule di apertura e chiusura della conversazione.	la comprensione di alcuni passaggi. Le strategie di compensazione, se presenti, sono spesso inefficaci.		
4	Tenta di realizzare i compiti assegnati ma non ci riesce o ci riesce solo in parte. Si esprime con una serie di enunciati giustapposti e incoerenti tra loro.	Nonostante gli sforzi dell'interlocutore l'interazione si blocca spesso. Ha difficoltà a segnalare di aver bisogno di ripetizioni e chiarimenti.	Repertorio insufficiente in rapporto alla consegna. A causa degli errori lessicali il messaggio è in gran parte incomprensibile.	Gamma ristrettissima di strutture. Errori frequenti impediscono la comprensione di parti importanti del discorso.	Malgrado gli sforzi dell'interlocutore la pronuncia scorretta rende incomprensibili parti importanti del discorso.
3	Frequenti interruzioni e blocchi ostacolano fortemente la comprensione.				
2	I compiti assegnati non sono portati a compimento. Si esprime con parole isolate.	I tentativi dell'interlocutore di stabilire un'interazione cadono nel vuoto. Non sa dire che non ha capito.	Repertorio limitatissimo. Gli errori lessicali rendono impossibile comprendere il messaggio.	Il candidato dà l'impressione di non possedere alcuna struttura grammaticale. Continui fraintendimenti da parte dell'interlocutore.	Malgrado gli sforzi dell'interlocutore, la pronuncia scorretta rende il discorso del tutto incomprensibile.
1					

Tab. 2. Griglia analitica di valutazione dell'interazione orale, livello A2 (Certificazione di italiano come lingua straniera PLIDA)

Apêndice 1. Tabelas analíticas para a avaliação da produção linguística

Apresentamos a seguir duas tabelas analíticas, uma em italiano e uma em português (constam em seu idioma original; não foram submetidas à tradução). A primeira, relatada a seguir, foi elaborada pela Secretaria Municipal de Educação do Rio de Janeiro (1º. bimestre de 2014), e tem como objetivo avaliar as habilidades de escrita dos alunos do 9º. ano do ensino fundamental (que corresponde à III média do sistema escolar italiano).

ASPECTOS	INDICADORES	PONTUAÇÃO
ESTRUTURA	<p>Apresentar elementos do texto, utilizando as características específicas da tipologia nos locais adequados e realizando a devida paragrafação.</p> <p>3 itens:</p> <ul style="list-style-type: none"> • Introdução; • Desenvolvimento; • Conclusão. 	<p>2,5 = Apresenta os 3 itens bem definidos, com uso adequado dos parágrafos.</p> <p>2,0 = Apresenta os 3 itens bem definidos, mas com falhas na paragrafação.</p> <p>1,5 = Apresenta os 3 itens, mas em um único parágrafo.</p> <p>1,0 = Apresenta 2 itens, mas com falhas na paragrafação.</p> <p>0,5 = Apresenta 1 ou dois itens, mas em um único parágrafo.</p> <p>0 = Não atende.</p>
COERÊNCIA	<p>Compreender a proposta de produção textual e desenvolver o tema, estabelecendo uma comunicação entre emissor e receptor por meio da elaboração de</p>	<p>2,5 = Narra a história, abordando os 4 itens, com coerência e clareza.</p> <p>2,0 = Narra a história, abordando os 4 itens, com algumas falhas na coerência ou na clareza.</p> <p>1,5 = Narra a história, abordando apenas 3 itens, porém escreve com coerência e clareza.</p>

	<p>um texto estruturado. 4 itens:</p> <ul style="list-style-type: none"> • Apresentação; • Conflito gerador; • Clímax; • Desfecho. 	<p>1,0 = Narra a história, abordando apenas 3 itens, porém escreve com falhas na coerência ou na clareza.</p> <p>0,5 = Narra a história, abordando apenas 1 ou 2 itens, com pouca consistência, apresentando falhas na coerência e na clareza.</p> <p>0 = Não consegue formar um texto.</p>
COESÃO	<p>Demonstrar conhecimento dos mecanismos linguísticos necessários para a construção da produção textual (Recursos coesivos: palavras ou expressões utilizadas para ligar as partes do texto ou para evitar repetições desnecessárias).</p>	<p>2,5 = Articula <u>adequadamente</u> as partes do texto, sem apresentar problemas na utilização de recursos coesivos.</p> <p>2,0 = Articula <u>adequadamente</u> as partes do texto, podendo apresentar eventuais problemas na utilização de recursos coesivos.</p> <p>1,5 = Articula <u>razoavelmente</u> as partes do texto, mas apresenta problemas na utilização de recursos coesivos.</p> <p>1,0 = Articula <u>razoavelmente</u> as partes do texto, mas apresenta uso excessivo de repetições desnecessárias.</p> <p>0,5 = Articula <u>precaricamente</u> as partes do texto, devido a problemas frequentes na utilização de recursos coesivos.</p> <p>0 = <u>Não</u> articula as partes do texto.</p>
USO DA LÍNGUA	<p>Utilizar a escrita como sistema de representação, respeitando os seguintes aspectos do código linguístico:</p> <ul style="list-style-type: none"> • uso adequado do vocabulário; • ortografia – relação 	<p>2,5 = Demonstra <u>excelente</u> domínio do código linguístico, respeitando os aspectos ortográficos, de pontuação e de concordância.</p> <p>2,0 = Demonstra <u>bom</u> domínio do código linguístico, respeitando os aspectos ortográficos, de pontuação e de concordância.</p> <p>1,5 = Demonstra <u>bom</u> domínio do código linguístico, respeitando os aspectos de concordância, mas com algumas dificuldades em pontuação ou acentuação.</p>

	<p>grafema/fonema, espaçamento entre palavras, uso de maiúsculas e acentuação;</p> <ul style="list-style-type: none"> • pontuação – ponto final, ponto de exclamação, ponto de interrogação, vírgula; <p>concordância nominal e verbal (relações morfossintáticas).</p>	<p>1,0 = Demonstra <u>bom</u> domínio do código linguístico, respeitando os aspectos ortográficos ou de pontuação, mas com dificuldades na concordância.</p> <p>0,5 = Demonstra conhecimento <u>reduzido</u> do código linguístico, apresentando dificuldades com relação à pontuação, à acentuação e à concordância.</p> <p>0 = Não apresenta conhecimento mínimo do código linguístico.</p>
--	--	--

Tab. 1. Tabela analítica de avaliação da produção escrita
(Secretaria Municipal da Educação, Prefeitura do Rio de Janeiro)¹⁵²

A segunda é utilizada na avaliação da interação oral na prova de proficiência de italiano como língua estrangeira PLIDA, no nível A2 (BARIVIERA *et al.*, 2021).

¹⁵² Per griglie di valutazione della produzione testuale elaborate dalla stessa istituzione, cfr. <<https://onedrive.live.com/view.aspx?resid=9A0409FEB0892781239&cid=09a0409feb089278&app=WordPdf>> (01.12.2021).

Para outras tabelas de avaliação da produção textual produzidas pela mesma instituição, cfr. <<https://onedrive.live.com/view.aspx?resid=9A0409FEB0892781239&cid=09a0409feb089278&app=WordPdf>> (01.12.2021).

	EFFICACIA COMUNICATIVA	INTERAZIONE	LESSICO	GRAMMATICA	PRONUNCIA
10	<p>Realizza i compiti assegnati in modo pienamente soddisfacente.</p> <p>Produce sequenze più lunghe, utilizzando semplici meccanismi di coesione.</p> <p>Nonostante esitazioni e false partenze, il discorso può essere seguito con uno sforzo contenuto.</p>	<p>Risponde agli stimoli dell'interlocutore in maniera adeguata.</p> <p>È in grado di chiedere ripetizioni e chiarimenti in modo molto semplice.</p> <p>Cerca di usare qualche formula di cortesia.</p>	<p>Repertorio pienamente soddisfacente in rapporto alla consegna.</p> <p>Gli errori lessicali possono essere frequenti, ma il messaggio è generalmente chiaro.</p> <p>Le strategie di compensazione, se presenti, sono efficaci.</p>	<p>Padroneggia le strutture previste per il livello.</p> <p>Nonostante la presenza di errori, il messaggio è ben comprensibile.</p>	<p>L'interlocutore fa poca fatica a interpretare i passaggi in cui la pronuncia è scorretta.</p>
9					
8	<p>Realizza i compiti assegnati in modo adeguato anche se alcuni punti possono essere sviluppati meno di altri.</p> <p>Formula enunciati brevi e semplici, con qualche forma di coesione.</p> <p>Esitazioni e false partenze sono frequenti e a tratti rendono il discorso faticoso da seguire.</p>	<p>Risponde agli stimoli dell'interlocutore in maniera quasi sempre adeguata.</p> <p>È in grado di chiedere ripetizioni e chiarimenti in modo molto semplice.</p> <p>Applica in modo adeguato le principali formule di apertura e chiusura della conversazione.</p>	<p>Repertorio adeguato alla consegna.</p> <p>Gli errori lessicali possono rendere alcuni passaggi poco chiari.</p> <p>Le strategie di compensazione, se presenti, non sono sempre efficaci.</p>	<p>Padroneggia abbastanza bene le strutture previste per il livello.</p> <p>Errori frequenti possono rendere la comprensione del messaggio occasionalmente difficoltosa.</p>	<p>In alcuni passaggi l'interlocutore fa fatica a interpretare il messaggio a causa della pronuncia scorretta.</p>
7					

6 5	<p>Realizza i compiti assegnati in parte o in modo approssimativo.</p> <p>Si esprime con una serie di enunciati giustapposti; il legame logico tra gli enunciati è comunque chiaro.</p> <p>Esitazioni e false partenze sono frequenti e rendono il discorso faticoso da seguire.</p>	<p>L'interazione procede grazie agli sforzi dell'interlocutore.</p> <p>È in grado di segnalare che ha bisogno di ripetizioni e chiarimenti in modo molto semplice anche non verbale.</p> <p>Applica con qualche incertezza le principali formule di apertura e chiusura della conversazione.</p>	<p>Repertorio appena sufficiente in rapporto alla consegna.</p> <p>Numerosi errori lessicali possono rendere faticosa la comprensione di alcuni passaggi.</p> <p>Le strategie di compensazione, se presenti, sono spesso inefficaci.</p>	<p>Padroneggia sufficientemente le strutture previste per il livello.</p> <p>La comprensione del messaggio può essere compromessa in più punti.</p>	<p>L'interlocutore fa costantemente fatica a interpretare il messaggio a causa della pronuncia scorretta.</p>
	<p>Tenta di realizzare i compiti assegnati ma non ci riesce o ci riesce solo in parte.</p> <p>Si esprime con una serie di enunciati giustapposti e incoerenti tra loro.</p> <p>Frequenti interruzioni e blocchi ostacolano fortemente la comprensione.</p>	<p>Nonostante gli sforzi dell'interlocutore l'interazione si blocca spesso.</p> <p>Ha difficoltà a segnalare di aver bisogno di ripetizioni e chiarimenti.</p>	<p>Repertorio insufficiente in rapporto alla consegna.</p> <p>A causa degli errori lessicali il messaggio è in gran parte incomprensibile.</p>	<p>Gamma ristrettissima di strutture.</p> <p>Errori frequenti impediscono la comprensione di parti importanti del discorso.</p>	<p>Malgrado gli sforzi dell'interlocutore la pronuncia scorretta rende incomprensibili parti importanti del discorso.</p>
4 3 2 1	<p>I compiti assegnati non sono portati a compimento.</p> <p>Si esprime con parole isolate.</p>	<p>I tentativi dell'interlocutore di stabilire un'interazione cadono nel vuoto.</p> <p>Non sa dire che non ha capito.</p>	<p>Repertorio limitatissimo.</p> <p>Gli errori lessicali rendono impossibile comprendere il messaggio.</p>	<p>Il candidato dà l'impressione di non possedere alcuna struttura grammaticale.</p> <p>Continui fraintendimenti da parte dell'interlocutore.</p>	<p>Malgrado gli sforzi dell'interlocutore, la pronuncia scorretta rende il discorso del tutto incomprensibile.</p>

Tab. 2. Tabela analítica de avaliação da interação oral, nível A2 (Certificação de italiano como língua estrangeira PLIDA)

Appendice 2. Griglia di analisi di un testo audio (livello B2)

Riportiamo una tabella atta ad accertare l'idoneità di un testo audio per la valutazione della comprensione orale di livello B2 di un apprendente adulto. Il valutatore, durante l'ascolto del brano, prende nota di una serie di aspetti e valuta se il brano può essere adatto, accertando in particolare che non sia troppo difficile (cioè, di livello C1). Domande che guidano la compilazione sono le seguenti:

- Il tema è troppo specialistico? È privo di riferimenti concreti?
- Vi è una sufficiente densità informativa, oppure le informazioni trasmesse sono scarse?
- L'articolazione del discorso è relativamente chiara, con una netta distinzione tra nuclei informativi e informazioni-satellite, oppure abbiamo una struttura estremamente complessa, con piste parallele, digressioni, analessi, ecc.?
- Il testo richiede, da parte del lettore, una serie di conoscenze molto specifiche? Sono presenti impliciti culturali?
- Abbiamo un lessico prevalentemente ad alta frequenza? O abbondano termini tecnici o formule idiomatiche poco comuni?
- La pronuncia è relativamente chiara, oppure ci sono numerosi tratti non standard?
- La durata dell'audio consente di mappare sufficientemente la competenza?
- La velocità dell'eloquio è normale o al massimo leggermente sostenuta (generando una impressione generale di chiarezza) oppure è caratterizzata da numerose accelerazioni?

- La qualità dell'audio è buona? Oppure ci sono elementi di disturbo (es. musica di sottofondo, rimbombo, ecc.)?
- Il brano è una unità testuale chiusa in sé o ci sono riferimenti ad altri testi (es. a interviste, ecc.), tali da opacizzare la comprensione?

TITOLO DEL BRANO	
LIVELLO IPOTETICO	
TEMA	
DENSITÀ INFORMATIVA	
ARTICOLAZIONE DEL DISCORSO	
CONOSCENZE RICHIESTE/ IMPLICITI	
LESSICO	
PRONUNCIA	
DURATA	
VELOCITÀ DI ELOQUIO	
QUALITÀ DELL' AUDIO	
RIFERIMENTI EXTRATESTUALI	

Tab. 3. Griglia di analisi di un testo audio (livello B2)

Apêndice 2. Tabela de análise de um texto de escuta (nível B2)

Apresentamos uma tabela para verificar a adequação de um texto de escuta, a fim de avaliar o nível de compreensão oral B2 de um aluno adulto. O avaliador, ao ouvir o texto, toma nota de uma série de aspectos e avalia se o texto de escuta pode ser adequado, assegurando, principalmente, que não seja difícil demais (ou seja, de nível C1). As perguntas que orientam o preenchimento são as seguintes:

- O assunto é muito específico? É desprovido de referências concretas?
- Há uma densidade informacional suficiente ou a informação transmitida é escassa?
- A estrutura do discurso é relativamente clara, com uma distinção clara entre os núcleos informativos e as informações secundárias ou temos uma estrutura extremamente complexa, com pistas paralelas, digressões, *flashbacks*, etc.?
- O texto exige do leitor uma série de conhecimentos muito específicos? Existem referências culturais implícitas?
- Temos, predominantemente, um léxico de alta frequência? Ou há um excesso de termos técnicos e expressões idiomáticas incomuns?
- A pronúncia é relativamente clara ou existem inúmeros trechos não-padrão?
- A duração do áudio permite que a competência seja suficientemente mapeada?

- A velocidade do discurso é normal ou, no máximo, ligeiramente mais rápida (dando uma impressão geral de clareza) ou é caracterizada por inúmeras acelerações?
- A qualidade do áudio é boa? Ou existem elementos de interferência (ex. música de fundo, eco, etc.)?
- O texto de escuta é uma unidade textual autônoma ou faz referência a outros textos (ex. entrevistas, etc.), de modo a embaçar a compreensão?

TÍTULO DO TEXTO	NÍVEL HIPÓTÉTICO	ASSUNTO	DENSIDADE INFORMACIONAL	ESTRUTURA DO DISCURSO	CONHECIMENTOS EXIGIDOS/ REFERÊNCIAS IMPLÍCITAS	LÉXICO	PRONÚNCIA	DURAÇÃO	VELOCIDADE DO DISCURSO	QUALIDADE DO ÁUDIO	REFERÊNCIAS EXTRATEXTUAIS

Tab. 3. Tabela de análise de um texto de escuta (nível B2)

Appendice 3. Griglia semplificata di valutazione di un testo argomentativo

Di seguito riportiamo una griglia di valutazione di un testo argomentativo (livello B2). Lo strumento è stato applicato durante un'esperienza di valutazione tra pari, all'interno di un corso di scrittura argomentativa in italiano come lingua straniera, presso l'Università Federale Fluminense (Niterói, Brasile). L'elaborato di ciascuno è stato valutato dai compagni, secondo la seguente scala di valutazione.

Segna da 1 a 5 il valore che attribuisce a ogni parametro. Puoi esprimere un commento

1= molto poco 2= poco 3= sufficiente 4= buono 5= molto buono

A ciascun valutatore era concesso esprimere un commento (in italiano o in L1), per suggerire all'autore come migliorare la propria *performance*.

	VOTO	COMMENTI
Il tema è chiaro		
C'è una introduzione adeguata		
C'è uno sviluppo adeguato		
C'è una conclusione adeguata		
Il testo è armonico e ben strutturato nel suo insieme		
Il testo è scritto in maniera da catturare l'interesse di chi legge		
L'autore è preciso nelle affermazioni		
Se esprime un giudizio, l'autore lo difende in maniera adeguata		
C'è una progressione delle idee; non si ripete sempre la stessa cosa		

Tab. 4. Griglia semplificata per la valutazione tra pari di un testo argomentativo

Apêndice 3. Tabela de avaliação simplificada de um texto argumentativo

Apresentamos, a seguir, uma tabela de avaliação de um texto argumentativo (nível B2). A tabela foi utilizada durante uma experiência de avaliação entre pares, no curso de redação argumentativa em italiano como língua estrangeira, na Universidade Federal Fluminense (Niterói, Brasil). O texto de cada um foi avaliado pelos colegas, de acordo com a seguinte escala de avaliação:

Atribua um valor de 1 a 5 para cada parâmetro. Você pode escrever um comentário

1= muito pouco 2= pouco 3= suficiente 4= bom 5= muito bom

Cada avaliador foi autorizado a comentar (em italiano ou em L1), sugerindo ao autor como melhorar a sua própria *performance*.

	NOTA	COMENTÁRIOS
O assunto é claro		
A introdução está adequada		
O desenvolvimento está adequado		
A conclusão está adequada		
O texto é harmônico e bem estruturado como um todo		
O texto é escrito para atrair o interesse do leitor		
O autor é preciso nas declarações		
Ao expor suas ideias, o autor sabe defendê-las adequadamente		
Há uma progressão de ideias; não se repete sempre a mesma coisa		

Tab. 4. Tabela simplificada de avaliação entre pares de um texto argumentativo

Appendice 4. Generatività. Approfondimento

Quale esempio di *prompt* poco generativo (che non incentiva cioè lo studente a esprimersi), si consideri il seguente, volto ad attestare una competenza orale di livello B2 (il testo originale inglese è accessibile al seguente *link* <<http://www.fetsystem.com/ieltssexamination/ielts-speaking-part-2-topic-the-embarrassing-moment-life>>).

IL MOMENTO PIÙ IMBARAZZANTE DELLA TUA VITA

- *Descrivi l'evento*
- *Perché ti sei sentito/a in imbarazzo?*
- *Come si possono evitare situazioni di questo genere?*

Descrivere a un estraneo, in un'altra lingua, il momento più imbarazzante della propria vita (al quale, dunque, sono associate emozioni tutt'altro che positive) è un compito di per sé imbarazzante, che non riflette una situazione spontanea.

Anche quest'altro esempio, citato in Van Lier (1989), dà adito a dubbi.

- *L'intervistatore assume il ruolo di un/a bimbo/a di 4 anni, il cui gattino è appena morto.*
- *Prova a consolare il/la bambino/a e ad offrire una soluzione al problema.*

Cosa si può dire, infatti, a un bimbo disperato che ha perso il suo animale preferito? O meglio, ha senso consolarlo a parole? Non vale forse più un tenero abbraccio?

Apêndice 4. Produtividade. Aprofundamento

O seguinte *prompt*, cujo objetivo é atestar um nível de competência oral B2, é pouco produtivo: não incentiva o estudante a se expressar (o texto original em inglês está acessível no seguinte link <<http://www.fetsystem.com/ielts/examination/ielts-speaking-part-2-topic/the-embarrassing-moment-life>>).

O MOMENTO MAIS CONSTRANGEDOR DA TUA VIDA

- *Descreva o ocorrido*
- *Por que você se sentiu envergonhado/a?*
- *Como podem ser evitadas tais situações?*

Descrever para um desconhecido, em outra língua, o momento mais constrangedor da própria vida (que, portanto, está associado a emoções nada positivas) é uma tarefa, por si só, embaraçosa, que não reflete uma situação espontânea.

Esse outro *prompt*, citado em Van Lier (1989), também suscita dúvidas.

- *O entrevistador assume o papel de uma criança de 4 anos, cujo gatinho acaba de morrer.*
- *Tente consolar a criança e oferecer uma solução para o problema.*

Na verdade, o que se pode dizer a uma criança desesperada que perdeu o seu animal favorito? Ou melhor, faz sentido consolá-la com palavras? Não vale muito mais um abraço carinhoso?

Appendice 5. Sottorappresentazione del costruito. Approfondimento

L'esercizio a seguire (attinto da AVELLA *et al.*, 2013: 98) è presentato come un'attività di scrittura rivolta a studenti di italiano di livello A2. In realtà, si tratta di un esercizio morfologico, con un minimo recupero lessicale; manca infatti la dimensione ideativa, di pianificazione e di organizzazione che un'attività di scrittura, anche a un livello basso, comporta. Si scambia per scrittura, così, un'esercitazione di tipo grammaticale (e in minima parte lessicale). La conoscenza delle strutture è in effetti una parte del costruito dello scrivere, non lo rappresenta del tutto.

Ieri hai avuto una giornata molto intensa; racconta a un amico cosa hai fatto, seguendo le informazioni scritte sul biglietto. Scrivi almeno 50 parole.

8.30 *Accompagnare Anna Asilo*

9 *Pane e verdure*

11 *Macelleria*

12 *Preparare pranzo*

16 *Telefonare a Luisa*

19 *Pizza*

Ciao Filippo,

oggi [sic] ho avuto molti impegni:

Apêndice 5. Sub-representação do constructo. Aprofundamento

O exercício a seguir (elaborado a partir de AVELLA *et al.*, 2013: 98) é apresentado como uma atividade de escrita para estudantes de italiano de nível A2. Na realidade, é um exercício morfológico, com um mínimo de reemprego de vocabulário; falta-lhe, de fato, a dimensão criativa, de planejamento e de organização que uma atividade de escrita requer, mesmo em um nível elementar. Confunde-se, então, a escrita com um exercício de tipo gramatical (e, em menor grau, lexical). O conhecimento das estruturas é, de fato, uma parte do constructo da escrita, não o representa completamente.

Ontem você teve um dia muito ocupado; diga a um amigo o que você fez, seguindo as informações anotadas. Escreva, no mínimo, 50 palavras.

8.30 *Acompanhar Anna Creche*

9 *Pão e feira*

11 *Açougue*

12 *Preparar o almoço*

16 *Telefonar para Luisa*

19 *Pizza*

Olá, Filippo!

Hoje [sic] eu tive muitos compromissos:

Appendice 6. Variabile esterna. Approfondimento

Quale esempio di invalidità dovuto a una variabile esterna al costrutto, diamo menzione all'attività "Re-tell lecture", inclusa nel pacchetto "Speaking" del *Pearson English Language Test Academic*. L'attività richiede al candidato di ascoltare un brano della durata di poco oltre 1 minuto, e di ripeterne il contenuto in un tempo massimo di 40 secondi (si veda una simulazione al seguente *link*: <<https://ptetutorials.com/sample-questions/speaking-re-tell-lecture>>).

Il costrutto di riferimento è quello della produzione orale ("speaking"), eppur tuttavia viene esercitata un'abilità integrata (frutto della combinazione tra comprensione orale e produzione orale). Peraltro, il fatto di far ripetere 'a caldo' un messaggio appena ascoltato (senza che sia concessa alcuna possibilità di rielaborazione cognitiva) non riflette un contesto naturale: è improbabile che un professore chieda dallo studente di ripetere cosa egli ha detto il minuto prima. L'idea che soggiace a questo tipo di esercitazione è che la memoria sia uno strumento di sintesi fedele e affidabile, mentre così non è: la memoria è uno strumento che rielabora e raffina in continuazione i dati in entrata (cfr. CARDONA, 2001). Abbiamo quindi (anche) un difetto di *autenticità*.

Apêndice 6. Variável externa. Aprofundamento

Como exemplo de falta de validade, devido a uma variável externa ao constructo, citamos a atividade “Re-tell lecture”, incluída no pacote “Speaking” do *Pearson English Language Test Academic*. A atividade exige que o candidato ouça uma gravação de pouco mais de 1 minuto, e que ele repita o conteúdo em 40 segundos, no máximo (veja uma simulação no seguinte link: <<https://ptetutorials.com/sample-questions/speaking-re-tell-lecture>>).

O constructo de referência é o da produção oral (“speaking”), no entanto, uma habilidade integrada é treinada (resultado da combinação entre compreensão oral e produção oral). Além disso, o fato de ter que repetir imediatamente uma mensagem que acaba de ser ouvida (sem qualquer possibilidade de reelaboração cognitiva) não reflete um contexto natural: é improvável que um professor peça ao estudante para repetir o que ele disse no minuto anterior. A ideia subjacente a este tipo de exercício é que a memória é uma ferramenta de síntese fiel e confiável, o que não é assim: a memória é um instrumento que processa e refina os dados recebidos continuamente (cfr. CARDONA, 2001). Portanto, temos também um defeito de *autenticidade*.

RIFERIMENTI BIBLIOGRAFICI

REFERÊNCIAS BIBLIOGRÁFICAS

- ACKERMAN, P. L.; KANFER, R., “Test Length and Cognitive Fatigue: An Empirical Examination of Effects on Performance and Test-taker Reactions”, *Journal of Experimental Psychology: Applied*, 15, 2, 2009, pp. 163-181.
- ALDERSON, C., “The Cloze Procedure and Proficiency in EFL”, *TESOL Quarterly*, 13, 1979, pp. 219-227.
- ALDERSON, C., *Assessing Reading*, CUP, Cambridge, 2000.
- ALDERSON, C.; BRUNFAUT, T.; HARDING, L., “Towards a Theory of Diagnosis in Second and Foreign Language Assessment: Insights from Professional Practice Across Diverse Fields”, *Applied Linguistics*, 36, 2, 2015, pp. 236-260.
- ALDERSON, C.; CLAPHAM, C.; Wall, D., *Language Test Construction and Evaluation*, CUP, Cambridge, 1995.
- ALLAN, D., “Testing and Assessment in ELT”, *English Teaching Professional*, 8, 1999.
- ALLEN, M. J.; YEN, W. M., *Introduction to Measurement Theory*, Brooks/Cole, Monterey, CA, 1978.
- ALTE, *Multilingual Glossary of Language Testing Terms*, CUP, Cambridge, 1999.
- ALTE, *A Manual for Language Test Development and Examining for Use with the CEFR*, Council of Europe, Strasbourg, 2011 <<https://www.coe.int/en/web/common-european-framework-reference-languages/documents>>.
- ANTIBI A., *La constante macabre*, Math’Ador, Toulouse, 2003.
- AVELLA, P.; BORGONOVO, P.; DESIDERIO, F., *Il test di italiano per stranieri*, Alpha Test, Milano, 2013.
- BACHMAN, L. F., *Fundamental Considerations in Language Testing*, OUP, Oxford, 1990.

- BACHMAN, L. F., *Statistical Analysis for Language Assessment*, CUP, Cambridge, 2004.
- BACHMAN, L. F., “What Is the Construct? The Dialectic of Abilities and Contexts in Defining Constructs in Language Assessment”, in J. FOX, M. WESCHE; D. BAYLISS; C. E. TURNER; C. DOE; C. LIYING (eds.), *Language Testing Reconsidered*, Les Presses de l’Université d’Ottawa, Ottawa, 2007, pp. 41-72.
- BAKER, J.; RINVOLUCRI, M., *Unlocking Self-expression through NLP*, Delta, Peaslake, 2005.
- BALBONI, P. E., *Didattica dell’italiano a stranieri*, Bonacci, Roma, 1994.
- BALBONI P. E., *Tecniche didattiche per l’educazione linguistica*, Utet, Torino, 1998.
- BALBONI P. E., *Fare educazione linguistica*, Utet, Torino, 2008.
- BANDURA, A., “Self-efficacy: Toward a Unifying Theory of Behavior Change”, *Psychological Review*, 84, 1977, pp. 191-215.
- BANDURA, A., *Self-efficacy: The Exercise of Control*, Freeman, New York, 1997.
- BARIVIERA, F.; CARDILLO, G.; Di TOMASSI, A.; MENZINGER, C.; VECCHIO, P., *Livello A2 PLIDA. Quaderno delle Specifiche. II edizione*, Società Dante Alighieri, Roma, 2021.
- BARNI, M., “Etica e politica della valutazione”, in E. LUGARINI (ed.), *Valutare le competenze linguistiche*, Franco Angeli, Milano, 2010, pp. 86-99.
- BENNETT B., “Assessment and Cooperative Learning: The Missing Think”, in S. SCOTT; D. SCOTT; C. WEBBER (eds.), *Leadership of Assessment, Inclusion, and Learning. The Enabling Power of Assessment*, Vol. 3, Springer, New York, pp. 45-85.
- BENVENUTO, G., *Mettere i voti a scuola: introduzione alla docimologia*, Carocci, Roma, 2003.
- BERNAUS, M., WILSON, A., “Teachers’ Motivation, Classroom Strategy Use, Students’ Motivation and Second Language Achievement”, *Porta Linguarum*, 12, 2009, pp. 25-36.

- BLACK, P.; WILLIAM, D., "Inside the Black Box: Raising Standards through Classroom Assessment", *Phi Delta Kappan*, 80, 1998, pp. 1-13. <<https://pdfs.semanticscholar.org/15bc/cadd19dbeb64ee5f0edac90e5857e6d5ad66.pdf>>
- BOEKAERTS, M.; PINTRICH, P.; ZEIDNER, M., (eds.), *Handbook of Self-Regulation*, Academic Press, San Diego, CA, 2000, pp. 239-268.
- BOND, T. G.; FOX, C. M., *Applying the Rasch Model: Fundamental Measurement in the Human Science*, Lawrence Erlbaum Associates, Mahwah, NJ, 2007.
- BORKOWSKI, J. G.; MATHUKRISHNA, N., "Moving Metacognition into the Classroom: «Working Models» and Effective Strategy Teaching", in M. PRESSLEY; K. R. HARRIS; J. T. GUTHRIE (eds), *Promoting Academic Competence and Literacy in Schools*, Academic Press, New York, 1992, pp. 477-501.
- BORSBOOM, D.; MARKUS, K. A., "Truth and Evidence in Validity Theory", *Journal of Educational Measurement*, 50, 1, 2013, pp. 110-114.
- BRINDLEY, G., "Assessing Listening Abilities", *Annual Review of Applied Linguistics*, 18, 1998, pp. 171-191.
- BRITISH COUNCIL, *Language Assessment in the Classroom*, MOOC Future Learn, 2020, <www.futurelearn.com> [accesso con password].
- BRITTON, J. N.; MARTIN, N. C.; ROSEN, H., "Multiple Marking of English Compositions: An Account of an Experiment", *Schools Council Examinations Bulletin*, 12, 1966, pp. 49-50.
- BUCK, G., *Assessing Listening*, CUP, Cambridge, 2001.
- CADAMURO, A., *Stili cognitivi e stili di apprendimento*, Carocci, Roma, 2004.
- CAMILLI, G.; SHEPARD, L., *Methods for Identifying Biased Test Items*, Sage, Thousand Oaks, CA, 1994.
- CANTOIA, M.; CARRUBA C.; COLOMBO, B., *Apprendere con stile. Metacognizione e strategie cognitive*, Carocci, Roma 2004.
- CARDONA, M., *Il ruolo della memoria nell'apprendimento delle lingue*, Utet, Torino, 2001.
- CARR, N., *Designing and Analyzing Language Tests*, OUP, Oxford, 2011.

- CARROLL, J.; CARTON, A.; WILDS, C., *An Investigation of Cloze Items in the Measurement of Achievement in Foreign Language. Laboratory for Research in Instruction*, Graduate School of Education, Harvard University, Cambridge, MA, 1959.
- CARSON, J. G.; NELSON, G. L., “Chinese Students’ Perceptions of ESL Peer Response Group Interaction”, *Journal of Second Language Writing*, 5, 1, 1996, pp. 1-19.
- CASAL, S., “Cooperative Assessment for Learning in CLIL Contexts”, *Estudios sobre Educación*, 31, 2016, pp. 159-157.
- CASSANY, D., *Reparar la escritura. Didáctica de la corrección de lo escrito*, Graó, Barcelona, 1993.
- CERVINI, C.; MARTARI, Y., “Soggettività, vaghezza e ambiguità nel *language assessment*. Note sui confini dell’attribuzione di valore”, in E. MAGNI; Y. MARTARI (eds.), *L’ambiguità nelle e tra le lingue. Prospettive a confronto*, Edizioni Dell’Orso, Alessandria, 2020, pp. 101–130.
- CHAPELLE, C., “Construct Definition and Validity Inquiry in SLA Research”, in L. F. BACHMAN; A. D. COHEN (eds.), *Interfaces between Second Language Acquisition and Language Testing Research*, CUP, Cambridge, 1998, pp. 32-70.
- CHAPELLE, C.; ABRAHAM, R., G., “Cloze Method: What Difference Does It Make?”, *Language Testing*, 7, 2, 1990, pp. 121-146.
- CHEN, Z.; HENNING, G., “Linguistic and Cultural Bias in Language Proficiency Tests”, *Language Testing*, 2, 2, 1985, pp. 155-163.
- CHIARI, I., “La procedura cloze, la ridondanza e la valutazione della competenza della lingua italiana”, *Italica*, 79, 4, 2002, pp. 525-540.
- CIZEK, J. G.; BUNCH, M. B., *Standard Setting. A Guide to Establishing and Evaluating Performance Standards on Tests*, Sage, Thousand Oaks, CA, 2006.

- COHEN, A. D. “The Coming of Age of Research on Test-taking Strategies”, *Language Assessment Quarterly*, 3, 4, 2006, pp. 307-330.
- COHEN, J. *Statistical Power Analysis for the Behavioural Sciences*, Prentice Hall, Harlow, 1988².
- CONNOR, U.; ASENAVAGE, K., “Peer Response Groups in ESL Writing Classes: How Much Impact on Revision?”, *Journal of Second Language Writing*, 3, 3, 1994, pp. 257-276.
- COOMBE, C., *An A to Z of Second Language Assessment: How Language Teachers Understand Assessment Concepts*, British Council, London, 2018
<www.britishcouncil.org/sites/default/files/a_to_z_glossary_final.pdf>.
- COUNCIL OF EUROPE, *European Language Portfolio (ELP). Principles and Guidelines*, Strasbourg, 2000.
- COUNCIL OF EUROPE, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Strasbourg, 2001.
- COUNCIL OF EUROPE, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment Companion Volume with New Descriptors*, Strasbourg, 2020 <<https://rm.coe.int/common-european-framework-of-reference-for-languages-learning-teaching/16809ea0d4>>
Edizione italiana, nella rivista *Italiano Lingua Due*, 12, 2, 2020, pp. 1-283.
<<https://riviste.unimi.it/index.php/promoitals/article/view/15120>>.
- CUMMING, A., “Assessing Integrated Skills”, in A. J. KUNNAN (ed.), *The Companion to Language Assessment*, Wiley, Hoboken, NJ, 2014, pp. 1-14.
- CUSHING WEIGLE, S. C., “Effects in Training on Raters of ESL Composition”, *Language Testing*, 11, 1994, pp. 197-223.
- CUSHING WEIGLE, S., *Assessing Writing*, CUP, Cambridge, 2002.
- DALOISO, M., *L'educazione linguistica dell'allievo con bisogni specifici. italiano, lingue straniere e lingue classiche*, Utet, Torino, 2015.

- DANCEY, C. P.; REIDY, J., *Statistics without Maths for Psychology*, Prentice Hall, Harlow, 2007.
- DAVIDSON, F.; LYNCH, B. K., *Testcraft: A Teacher's Guide to Writing and Using Test Specifications*, Yale University Press, New Haven, CT, 2001.
- DAVIES, A.; BROWN, A.; ELDER, C.; HILL, K.; LUMLEY, T.; McNAMARA, T., *Dictionary of Language Testing*, CUP, Cambridge, 1999.
- DAVIS, P.; RINVOLUCRI, M., "Exercises that Help to Build Up Confidence During Exam Time", *Humanising Language Teaching Journal*, 8, 5, 2006 <<https://old.hltnmag.co.uk/sep06/ex.htm#C8>>.
- DE BENI, R.; MOÈ, A., *Motivazione e apprendimento*, Il Mulino, Bologna, 2000.
- D'ESTE, C., "New Views of Validity in Language Testing", *ELLE*, 1, 1, 2012, pp. 61-76.
- DÖRNYEI, Z., *Research Methods in Applied Linguistics*, CUP, Cambridge, 2007.
- DÖRNYEI, Z.; USHIODA, E., *Teaching and Researching Motivation*, Routledge, London, 2021³.
- DOWNING, S. M., "Twelve Steps for Effective Test Development", in S. M. DOWNING; T. M. HALADYNA (eds.), *Handbook of Test Development*, Lawrence Erlbaum Associates Publishers, Mahwah, NJ, 2006, pp. 3-25.
- ECKES, T., "Examining Rater Effects in TestDaF Writing and Speaking Performance Assessments: A Many-facet Rasch analysis", *Language Assessment Quarterly*, 2, 2005, pp. 197-221.
- ECKES, T., *Introduction to Many-Facet Rasch Measurement: Analyzing and Evaluating Rater-Mediated Assessments*, Peter Lang, Frankfurt am Main, 2015².
- ELIAS, M. J., HUNTER, L., KRESS, J. S., "Emotional Intelligence and Education", in J. CIARROCHI; J. P. FORGAS; J. D. MAYER (eds.), *Emotional Intelligence in Everyday Life: A Scientific Inquiry*, Psychology Press, Philadelphia, 2001, pp. 133-149.

- FAN, J.; McNAMARA, T.; KNOCH, U., *Fairness, Justice and Language Assessment*, OUP, Oxford, 2019.
- FIGUERAS, N.; NOIJONS, J. (eds.), *Linking to the CEFR Levels: Research Perspectives*, EALTA Research Colloquium, 2009 <www.ealta.eu.org/documents/resources/Research_Colloquium_report.pdf>
- FRAILE, J., PANADERO, E.; PARDO, R., “Co-creating Rubrics: The Effects on Self-regulated Learning, Self-efficacy and Performance of Establishing Assessment Criteria with Students”, *Studies in Educational Evaluation*, 53, 2017, pp. 69-76.
- FREIRE, P., *Pedagogia do Oprimido*, Paz e Terra, São Paulo, 1987.
- FULCHER, G., *Re-examining Language Testing. A Philosophical and Social Inquiry*, Routledge, London, 2015.
- FULCHER, G.; DAVIDSON, F., *Language Testing and Assessment: An Advanced Resource Book*, Routledge, London, 2007.
- FULCHER, G.; DAVIDSON, F.; KEMP, J., “Effective Rating Scale Development for Speaking Tests: Performance Decision Trees”, *Language Testing*, 28, 1, 2010, pp. 5-29.
- GALTON, F., “Vox Populi”, *Nature*, 75, 1907, pp. 450-451.
- GARDNER, H., FELDMAN, D. H.; KRECHEVSKY M., *Project Spectrum: Preschool Assessment Handbook*, Teachers College Press, New York, 1998.
- GARDNER, R. C., *Social Psychology and Language Learning: The Role of Attitudes and Motivation*, Edward Arnold, London, 1985.
- GAZZANIGA, M., *The Social Brain*, Basic Books, New York, 1985.
- GERANPAYEH, A., “Using DIF to Explore Item Difficulty in CAE Listening”, *Cambridge ESOL Research Notes*, 32, 2008, pp. 16-23.
- GERANPAYEH, A.; KUNNAN, A. J., “Differential Item Functioning in Terms of Age in the Certificate in Advanced English Examination”, *Language Assessment Quarterly*, 4, 2, 2007, pp 190-222.

- GIPPS, C., *Beyond Testing*, Falmer, Brighton, 1994.
- GRABOWSKI, K. C.; DAKIN, J. W., “Test Development Literacy”, in A. J. KUNNAN (ed.), *The Companion to Language Assessment*, Wiley, Hoboken, NJ, 2014, pp. 751-768.
- GREEN, R., *Statistical Analyses for Language Teachers*, Palgrave, London, 2013.
- GREEN, R., *Designing Listening Tests: A Practical Approach*, Palgrave, London, 2017.
- HAMP-LYONS, L., “Ethics in Language Testing”, in C. CLAPHAM (ed.), *The Encyclopedia of Language and Education*, vol. 7, Kluwer, Dordrecht, 1998, pp.323-333.
- HANSEN, J.; STANSFIELD, C., “The Relationship of Field Dependent-independent Cognitive Styles to Foreign Language Achievement”, *Language Learning*, 31, 1, 1981, pp. 349-367.
- HANSEN, J.; STANSFIELD, C., “Student Teacher Cognitive Styles and Foreign Language Achievement: A Preliminary Study”, *Modern Language Journal*, 66, 1982, pp. 263-273.
- HANSEN, J.; STANSFIELD, C., “Field Dependence-Independence as a Variable in Second Language Cloze Test Performance”, *TESOL Quarterly*, 17, 1, 1983, pp. 29-38.
- HARTER, S., “Effectance Motivation Reconsidered: Toward a Developmental Model”, *Human Development*, 21, 1978, pp. 34-64.
- HAUKÅS, A.; BJØRKE, C.; DYPEDAHL, M., (eds.), *Metacognition in Language Learning and Teaching*, Routledge, London, 2018.
- HOLLAND, P. W.; WAINER, H. (eds.), *Differential Item Functioning*, Lawrence Erlbaum, Hillsdale, NJ, 1993.
- HOUSE, J., “Translation Quality Assessment: Past and Present”, in Id. (ed.), *Translation: A Multidisciplinary Approach*, Palgrave, London, 2013, pp. 241-264.

- JAMIESON, J., "Defining Constructs and Assessment Design", in A. J. KUNNAN (ed.), *The Companion to Language Assessment*, Wiley, Hoboken, 2013, pp. 1-16.
- JOHNSON, D.; JOHNSON, R., "Cooperative Learning and Assessment", in D. KLUGE; S. McGUIRE; D. JOHNSON; R. JOHNSON (eds.), *JALT Applied Materials: Cooperative Learning*, Japan Association for Language Teaching, Tokyo, 1999, pp. 164-178.
- KANE, M. T., *An Argument-based Approach to Validation*, The American College Testing Program, Iowa City, 1990.
- KANE, M. T., "An Argument-based Approach to Validity", *Psychological Bulletin*, 112, 3, 1992, pp. 527-535.
- KANE, M. T., "Current Concerns in Validity Theory", *Journal of Educational Measurement*, 38, 4, 2001, pp. 319-342.
- KARAMI, H.; SALMANI NODOUSHAN, M. A., "Differential Item Functioning (DIF): Current Problems and Future Directions", *International Journal of Language Studies*, 5, 3, 2011, pp. 133-142.
- KATZ, S.; LAUTENSCHLAGER, G. J., "Answering Reading Comprehension Items Without Passages on The SAT-I, The ACT, and The GRE", *Educational Assessment*, 2, 4, 1994, pp. 295-308.
- KIM, M., "Detecting DIF Across the Different Language Groups in a Speaking Test", *Language Testing*, 18, 1, 2001, pp. 89-114.
- KOHN, A., *Punished by Rewards*, Mariner Books, New York, 1993.
- KUHL, J., "A Functional-Design Approach to Motivation and Self-Regulation", in M. BOEKAERTS; P. PINTRICH; M. ZEIDNER (eds.), *Handbook of Self-Regulation*, Academic Press, San Diego, CA, 2000, pp. 239-268.
- KUNNAN, A. J., "Differential Item Functioning and Native Language and Gender Groups: The Case of an ESL Placement Examination", *TESOL Quarterly*, 24, 1990, pp. 741-746.

- KUNNAN, A. J.; CARR, N. T., “Statistical Analyses of Test Results”, in C. A. CHAPPELLE (ed.), *The Encyclopedia of Applied Linguistics*, Vol. 2, Blackwell, Oxford, 2012, pp. 5396-5403.
- LADO, R., *Language Testing*, Longman, London, 1961.
- LANTOLF, J. P.; POEHNER, M. E., “Dynamic Assessment in the Classroom: Vygotskian Praxis for Second Language Development”, *Language Teaching Research*, 15, 1, 2010, pp. 11-33.
- LEAMNISON, R. N., *Thinking about Teaching and Learning: Developing Habits of Learning with First Year College and University Students*, Stylus, Sterling, VA, 1999.
- LEUNG, C., “Dynamic Assessment: Assessment for and as Teaching”, *Language Assessment Quarterly*, 4, 3, 2007, pp. 257-278.
- LIM, G. S., “The Development and Maintenance of Rating Quality in Performance Writing Assessment: A Longitudinal Study of New and Experienced Raters”, *Language Testing*, 28, 2011, pp. 543-560.
- LINACRE, J. M., *Facets Rasch Model Computer Program* [Software manual], Chicago: Winsteps.com, 2008.
- LITTLE, D.; PERCLOVA, R., *The European Language Portfolio: A Guide for Teachers and Teacher-Trainers*, Council of Europe, Strasbourg, 2001.
- LOUMA, S. (2004), *Assessing Speaking*, CUP, Cambridge.
- LUMLEY, T., “Assessment Criteria in a Large-scale Writing Test: What do They Really Mean to Raters?”, *Language Testing*, 19, 2002, pp. 246-276.
- MACARO, E., “The Relationship between Strategic Behaviour and Language Learning Success”, in E. MACARO (ed.), *The Continuum Companion to Second Language Acquisition*, Continuum, London, 2010, pp. 268-299.
- MACARO, E.; GRAHAM, S.; WOORE, R., *Improving Foreign Language Teaching*, Routledge, London, 2016.

- MACCIÒ, M., “Scopi della verifica e ripercussioni didattiche: che cosa rivela il washback di un test? Il caso del Victorian Certificate of Education”, *SILTA*, 3, 2017, pp. 549-568.
- MADSEN, H. S., “Determining the Debilitative Impact of Text Anxiety”, *Language Learning*, 32, 1, 1982, pp. 133-143.
- MAGRIS, M., “La valutazione della qualità della traduzione nella teoria e nella pratica”, in G. BENELLI; G. TONINI (eds.), *Studi in ricordo di Carmen Sánchez Montero*, EUT Edizioni Università di Trieste, Trieste, Vol. 1, 2006, pp. 183-194.
- MAHAL GALTI, A.; SAIDU, S.; YUSUF, H.; ALKANI GONI, A., “Rating Scale in Writing Assessment: Holistic vs. Analytical Scales: A Review”, *International Journal of English Research*, 4, 6, 2018, pp. 4-6.
- MANSELL, E., *Education by Numbers. The Tyranny of Testing*, Methuen, London, 2007.
- MARIANI, L.; POZZO G., *Stili, strategie e strumenti nell'apprendimento linguistico*, La Nuova Italia, Firenze, 2002.
- MARIANI, L.; TOMAI P., *Il Portfolio delle lingue*, Carocci, Roma, 2004.
- MATSUNO, S., “Self-, Peer- and Teacher- Assessment in Japanese University EFL Classrooms”, *Language Testing*, 26, 1, 2009, pp. 75-100.
- MAY, L., “Co-constructed Interaction in a Paired Speaking Test: The Rater’s Perspective”, *Language Testing*, 26, 3, 2009, pp. 397-421.
- McDONOUGH, S. K. M., “Promoting Self-Regulation in Foreign Language Learners”, *The Clearing House*, 74, 6, 2001, pp. 323-326.
- McNAMARA, T., *Measuring Second Language Performance*, Longman, London, 1996.
- MENDOZA RAMOS, A., “El uso de Many-Facet Rasch Measurement para examinar la calidad del proceso de corrección de pruebas de desempeño”, *Revista mexicana de investigación educativa*, 23, 77, 2018, pp. 597-625.
- MESSICK, S., “Validity”, in R. L. LINN (ed.), *Educational Measurement*, American Council on Education/Macmillan, New York, 1989, pp. 13-104.

- MURPHEY, T., *Language Hungry!*, Helbling, Innsbruck, 2006.
- MYFORD, C. M.; WOLFE, E. W., “Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part I”, *Journal of Applied Measurement*, 4, 2003, pp. 386-422.
- MYFORD, C. M.; WOLFE, E. W., “Detecting and Measuring Rater Effects Using Many-Facet Rasch Measurement: Part II”, *Journal of Applied Measurement*, 5, 2004, pp. 189-227.
- NJIEYEP NJOUOGWA, N. L., *Osservare l'interlingua a scuola raccolta e analisi dei dati*, Tesi di Laurea, Università di Modena-Reggio Emilia, Modena, 2007.
- NOVELLO, A., *Valutare una lingua straniera. Le certificazioni europee*, Cafoscarina, Venezia, 2009.
- NOVELLO, A., “Motivare alla valutazione linguistica”, *ELLE*, 1, 1, 2012, pp. 91-110.
- NOVELLO, A., *La valutazione delle lingue straniere e seconde nella scuola: Dalla teoria alla pratica*, Ca' Foscari, Venezia, 2014
<<https://edizionicafoscarini.unive.it/media/pdf/books/978-88-97735-90-8/978-88-97735-90-8.pdf>>.
- NOVELLO, A., “Gli studenti *gifted*: riflessioni e proposte per la loro valutazione linguistica”, *ELLE*, 7, 3, 2018, pp. 392-412.
- NOVELLO A., “La valutazione delle competenze linguistiche nella didattica a distanza”, *AggiornaMenti*, 18, 11, 2021, pp. 18-22
<<https://adi-germania.org/it/aggiornamenti-18/>>.
- OACKEY, G. J.; WAGNER, W. (eds.), *Assessing LS Listening. Moving towards Authenticity*, John Benjamins, Amsterdam, 2018.
- OLIVERI, M. E.; LAWLESS, R.; YOUNG, J. W., *A Validity Framework for the Use and Development of Exported Assessments*, ETS, 2005 <https://www.ets.org/s/about/pdf/exported_assessments.pdf>.

- ONAIBA, A.; JANNAT, F., “Test Method Effect and Test-Takers’ Scores: A Critical Review of the Pertinent Literature”, *Scientific Journal of Faculty of Education, Misurata University-Libya*, 1, 14, 2019, pp. 3-22.
- O’SULLIVAN, B., “Assessing Speaking”, in A. J. KUNNAN (ed.), *The Companion to Language Assessment*, Wiley, Hoboken, NJ, 2013, pp. 156-171.
- OXFORD, R. L., *Language Learning Strategies: What Every Teacher Should Know*, Newbury, New York, 1990.
- PAE, T.-I., “DIF for Examinees with Different Academic Backgrounds”, *Language Testing*, 21, 1, 2004, pp. 53-73.
- PALLOTTI, G., “Le ricadute didattiche delle ricerche sull’interlingua”, in E. JAFRANCESCO (ed.), *L’acquisizione dell’italiano L2 da parte di immigrati adulti*, Edilingua, Atene, 2005, pp. 43-59.
- PANADERO, E., “Implementing Summative Assessment with a Formative Flavour: A Case Study in A Large Class”, *Assessment & Evaluation in Higher Education*, 43, 2, 2018, pp. 307-322.
- PAULETTO, F., “L’analisi della conversazione per valutare l’autenticità dei materiali audio per l’insegnamento dell’italiano L2: una proposta metodologica”, *Euro.American Journal of Applied Linguistics and Languages*, 7, 2, 2020, pp. 28-50
- PELLEREY, M., *Le competenze individuali e il portfolio*, La Nuova Italia, Firenze, 2004.
- PENNINGTON, M.; ROGERSON-REVELL, P., “Assessing Pronunciation”, in M. PENNINGTON; P. ROGERSON-REVELL (eds.), *English Pronunciation. Teaching and Research. Contemporary Perspectives*, Palgrave, London, 2019, pp. 287-342.
- POLLIT, A.; MURRAY, N. L., “What Raters really Pay Attention to”, in M. MILANOVIC; N. SAVILLE (eds.), *Performance Testing, Cognition and Assessment*, CUP, Cambridge, 1996, pp. 1-7.

- PORTER, D., “The Effects of Quantity of Context on the Ability to Make Linguistic Predictions: A Flow in a Measure of General Proficiency”, in A. HUGHES; D. PORTER (eds.), *Current Developments in Language Testing*, Academic Press, London, 1983, pp. 63-74.
- POZZOLO, P., *Traduttore statistiche-italiano*, 2020 [e-book], <<https://paolapozzolo.it/risorse-gratuite>>
- POWERS, D. E.; WILSON, S. T., “Passage Dependence on the New SAT Reading Comprehension Questions”, *College Board Report*, 93, 3, 1993, pp. 1-18.
- PUNCH, K. F., *Introduction to Social Research*, Sage, Thousand Oaks, CA, 2005².
- PURPURA, J. E., *Assessing Grammar*, CUP, Cambridge, 2004.
- RASTELLI, S., “Riflettere sulla lingua serve ad impararla? Dipende da cosa si intende per «lingua» e «imparare»”, in S. RASTELLI; C. BAGNA (eds.), *Manifesto programmatico per l'insegnamento della lingua italiana agli studenti internazionali. Otto commenti*, Pacini, Pisa, 2020, pp. 49-72.
- REA-DICKINS P., “Currents and Eddies in the Discourse of Assessment: A Learning-focused Interpretation”, *International Journal of Applied Linguistics*, 16, 2, 2008, pp. 163-188.
- READ, G., *Assessing Vocabulary*, CUP, Cambridge, 2000.
- ROCCA, L., “Controllo, verifica e valutazione,” in A. BORRI; F. MINUZ; L. ROCCA (eds.), *Progettare percorsi di L2 per adulti stranieri. Dall'alfabetizzazione all'A1*, Loescher, Torino, 2016, pp. 90-107.
- ROEVER, C., “DIF in the Assessment of Second Language Pragmatics”, *Language Assessment Quarterly*, 4, 2, 2007, pp. 165-189.
- ROSENTHAL, R.; JACOBSON, L., *Pigmalione in classe*, Franco Angeli, Milano, 1977.
- RYAN, K. E.; BACHMAN, L. F., “Differential Item Functioning on Two Tests of EFL Proficiency”, *Language Testing*, 9, 1, 1992, pp. 12-29.
- SARIG, G., “Testing Meaning Construction: Can We Do It Fairly?”, *Language Testing*, 6, 1, 1989, pp. 77-94.

- SASAKI, M., “A Comparison of Two Methods for Detecting DIF in an ESL Placement Test”, *Language Testing*, 8, 2, 1991, pp. 95-111.
- SCARAMUCCI, M., “Proficiência em LE: considerações terminológicas e conceituais”, *Trabalhos em Linguística Aplicada*, 36, 2000, pp. 11-22.
- SCARAMUCCI, M., “Efeito retroativo da avaliação no ensino/aprendizagem de línguas: o estado da arte”, *Trabalhos em Linguística Aplicada*, 43, 2, 2004, pp. 203-226.
- SCARAMUCCI, M., “O exame Celpe-Bras e a proficiência do professor de português para falantes de outras línguas”, *Revista Digilengua*, 12, 2012, pp. 48-67.
- SCHEDL, M.; MALLOY, J., “Writing Items and Tasks”, in A. J. KUNNAN (ed.), *The Companion to Language Assessment*, Wiley, Hoboken, NJ, 2004, pp. 1-18.
- SCHOFFEN, J., MENDEL, K. “As especificações do exame Celpe-Bras e a descrição das tarefas da parte escrita: convergências e divergências”, *Domínios de Linguagem*, 12, 2, 2018, pp. 1091-1122.
- SHAW, S. D.; WEIR, C. J., *Examining Writing. Research and Practice in Assessing Second Language Writing*, CUP, Cambridge, 2007.
- SHOHAMY, E., “Does the Testing Method Make a Difference? The Case of Reading Comprehension”, *Language Testing*, 1, 2, 1984, pp. 147-170.
- SHOHAMY, E., “Testing Methods, Testing Consequences: Are They Ethical? Are They Fair?”, *Language Testing*, 14, 3, 1997, pp. 340-349.
- SHOHAMY, E., *The Power of Tests: A Critical Perspective on the Use of Language Tests*, Routledge, London, 2001.
- SHOHAMY, E.; GORDON, C.; KRAEMER, R., “The Effect of Rater’s Background and Training on the Reliability of Direct Writing Task”, *Modern Language Journal*, 76, 1992, pp. 27-33.
- SHOHAMY, E.; INBAR, O., “Validation of Listening Comprehension Tests: The Effect of Text and Question Type”, *Language Testing*, 8, 1, 1991, pp. 41-66.

- SHOHAMY, E.; REVES, T., “Authentic Language Tests: Where from and Where to?”, *Language Testing*, 2, 1, 1985, pp. 48-59.
- SERRAGIOTTO, G., *La valutazione nel CLIL: format e griglie*, Cafoscarina, Venezia, 2006.
- SERRAGIOTTO, G., *La valutazione degli apprendimenti linguistici*, Loescher, Torino, 2016.
- SPINELLI, B., “Il «dialogo» tra insegnante e studente nella produzione, revisione e valutazione della scrittura in italiano L2”, *Officina.it*, 22, 2014 <www.almaedizioni.it>.
- SPINELLI, B.; PARIZZI, F. (eds.), *Profilo della lingua italiana. Livelli di riferimento del QCER A1, A2, B1, B2*, La Nuova Italia, Milano, 2010.
- TAI, J; AJJAWI, R.; BOUD, D.; DAWSON, P.; PANADERO, E., “Developing Evaluative Judgement: Enabling Students to Make Decisions about the Quality of Work”, *Higher Education*, 76, 2018, pp. 467–481.
- TARDIEU, C.; HILDÉN, R.; LEHMANN, M.; REICHERT, M., *The CEF-ESTIM Grid, Level of Estimation Grid for Teachers*, European Centre for Modern Languages, Strasburgo, 2010 <www.ecml.at/tabid/277/PublicationID/74/Default.aspx>.
- TESSARO, F., *I fondamenti della valutazione scolastica*, Laboratorio Università-Scuola RED-SSIS Veneto, Venezia, 2004. <<http://www.univirtual.it/red/files/file/Tessaro-FondamentiValutazione.pdf>>.
- TORRESAN, P., “Il dettato e il dictocloze sono prove valide per valutare la comprensione orale?”, *Romanica Cracoviensia*, 14, 2014, pp. 138-150.
- TORRESAN, P., *Studio su cloze mirati della certificazione CILS. Livello avanzato*, Dialogarts, Rio de Janeiro, 2015.
- TORRESAN, P., “Analisi della predittività in prove di comprensione della certificazione .IT per la fascia di competenza dell’autonomia”, *Trabalhos em Língua Aplicada*, 55, 1, 2016, pp. 155-183.

- TORRESAN, P., “Raccomandazioni per chi redige un test linguistico”, *Bollettino Itals*, 17, 80, 2019, pp. 94-107.
- TORRESAN, P., “Recomendações para quem redige uma prova linguística”, *Caderno Seminal Digital*, 36, 6, 2020a, 297-320.
- TORRESAN, P., “La mappatura del testo centrata sugli studenti”, *Italiano Lingua Due*, 12, 1, 2020b, pp. 940-967.
- TORRESAN, P., *Valutare la comprensione in italiano L2/LS. Griglie per l'analisi e la selezione di testi complessi (B2-C2)*, Dialogarts, Rio de Janeiro, 2020c.
- TORRESAN, P., *Certificazione PLIDA Nuovo formato. Costrutti operativi*, Dialogarts, Rio de Janeiro, 2022.
- TSAGARI, C., *Method Effect on Testing Reading Comprehension: How Far Can We Go?*, MA Thesis Dissertation, University of Lancaster, Lancaster, UK, 1994.
- UNDERHILL, N., *Testing Spoken Language: A Handbook of Oral Testing Techniques*, CUP, Cambridge, 1987.
- Van LIER, L., “Reeling, Writhing, Drawling, Stretching, and Fainting in Coils: Oral Proficiency Interviews as Conversation”, *TESOL Quarterly*, 23, 2, 1989, pp. 489-508.
- VEDOVELLI, M., *Manuale della certificazione dell'italiano L2*, Carocci, Roma, 2005.
- VICENTINI, M., “As tarefas de compreensão oral para produção escrita no Celp-Bras: construto e operacionalização”, *Estudos Linguísticos*, 48, 1, 2019, pp. 561-580.
- WALL, D.; ALDERSON, J. C., “Examining Washback: The Sri Lankan Impact Study”, *Language Testing*, 10, 1993, pp. 41-69.
- WEINER, B., “A Theory of Motivation for Some Classroom Experiences”, *Journal of Educational Psychology*, 71, 1979, pp. 3-25.
- WEIR, C. J., *Language Testing and Validation. An Evidence-based Approach*, CUP, Cambridge, 2005.

- WEIR, C.; VIDAKOVIĆ, I.; GALACZI, E. D. (eds.), *Measured Constructs. A History of Cambridge English Language Examinations 1913-2012*, CUP, Cambridge, 2013.
- WHITE, R. W., “Motivation Reconsidered: The Concept of Competence”, *Psychological Review*, 12, 1959, pp. 265-310.
- WILLIAMS, M., “A Holistic-Componential Model for Assessing Translation Student Performance and Competency”, *Mutatis Mutandis*, 6, 2, 2013, pp.419-443
- WOOD, R.; QUINN, B., “Double Impression Marking of English Language Essay and Summary Questions”, *Educational Review*, 28, 3, 1976, pp. 229-245.
- WRIGHT, B. D.; MASTERS, G. N., *Rating Scale Analysis. Rasch Measurement*, Mesa, Chicago, 1982.

QUESTO GLOSSARIO PERMETTE DI DISPORRE DI UNA MAPPA DEI TERMINI-CHIAVE DELL'EDUCAZIONE LINGUISTICA. STRUMENTO UTILE A INSEGNANTI E A TUTTI COLORO CHE SI INTERESSANO ALL'ARGOMENTO, IL GLOSSARIO SERVE SIA A ORIENTARE IL LETTORE CHE A STIMOLARLO VERSO SUCCESSIVI APPROFONDIMENTI.

ESTE GLOSSÁRIO FORNECE UM MAPA DOS PRINCIPAIS TERMOS DA EDUCAÇÃO LINGUÍSTICA. UMA FERRAMENTA ÚTIL PARA OS PROFESSORES E PARA TODOS AQUELES QUE ESTÃO INTERESSADOS NO ASSUNTO, O GLOSSÁRIO SERVE TANTO PARA ORIENTAR O LEITOR COMO PARA ESTIMULÁ-LO A APROFUNDAR O ESTUDO.

