

National Agency for the Evaluation of
Universities and Research Institutes



Agenzia Nazionale di Valutazione del
sistema Universitario e della Ricerca

Evaluation of Research Quality



Valutazione Qualità della Ricerca

Valutazione della Qualità della Ricerca 2004-2010 (VQR 2004-2010)

Rapporto finale di area
Gruppo di Esperti della Valutazione dell'Area 11 (GEV 11)



RAPPORTO FINALE DI AREA.....	1
GRUPPO DI ESPERTI DELLA VALUTAZIONE DELL'AREA 11 (GEV 11).....	1
LISTA TABELLE.....	3
1 INTRODUZIONE	22
1.1 DESCRIZIONE DELL'AREA	28
1.2 LA COMPOSIZIONE DEL GEV	30
1.3 I TEMPI.....	31
2 LA VALUTAZIONE DEI "PRODOTTI DI RICERCA"	32
2.1 I CRITERI DI VALUTAZIONE.....	32
2.2 LA CLASSIFICAZIONE DELLE RIVISTE	34
2.3 IL PROCESSO DI VALUTAZIONE	36
2.4 I RISULTATI DELLA VALUTAZIONE	43
2.5 I PRODOTTI CONFERITI E LA CLASSIFICAZIONE DELLE RIVISTE	44
3 LA VALUTAZIONE DI AREA DELLE STRUTTURE	46
3.1 GLI INDICATORI DI QUALITÀ DELLE STRUTTURE NELL'AREA	48
3.1.1 <i>Premessa</i>	48
3.1.2 <i>Il primo indicatore</i>	49
3.1.3 <i>Il secondo indicatore</i>	49
3.1.4 <i>Ranking delle strutture sulla base dei due indicatori</i>	50
3.2 L'INDICATORE $IRAS1_{i,j}$ DEL BANDO VQR	51
3.3 COMMENTI SUL SIGNIFICATO DEGLI INDICATORI DI STRUTTURA DI AREA	52
4 LA VALUTAZIONE DI AREA DEI DIPARTIMENTI.....	53
4.1 PREMESSA	53
4.2 GLI INDICATORI DI QUALITÀ DI AREA DEL DIPARTIMENTO.....	54
4.2.1 <i>Il primo indicatore</i>	54
4.2.2 <i>Il secondo indicatore</i>	54
4.2.3 <i>Graduatorie dei Dipartimenti sulla base dei due indicatori</i>	55
4.3 L'INDICATORE $IRD1_{i,j,k}$ DEL BANDO VQR	56
5 ANALISI DEI RISULTATI.....	57
6 CONCLUSIONI.....	63



LISTA TABELLE

Tabella 1.1 - Delimitazione area 11 (per SSD)

Tabella 1.2 - Distribuzione dei prodotti per tipologia e SSD, Area 11 - sub-area non bibliometrica (Antropologia, Filosofia, Geografia, Pedagogia, Scienze del Libro e del Documento, Storia)

Tabella 1.3 - Distribuzione dei prodotti per tipologia e SSD, Area 11 - sub-area bibliometrica (Scienze Motorie, Psicologia)

Tabella 1.4 - Distribuzione dei prodotti per lingua e SSD, Area 11 - sub-area non bibliometrica

Tabella 1.5 - Distribuzione dei prodotti per lingua e SSD, Area 11 - sub-area bibliometrica

Tabella 1.6 - Soggetti valutati di area 11, per subGEV

Tabella 1.7 - Soggetti valutati di area 11, per tipo di struttura di afferenza

Tabella 1.8 - Numero di Soggetti Valutati (SV) e relativo numero di prodotti attesi (subarea non bibliometrica)

Tabella 1.9 - Numero di Soggetti Valutati (SV) e relativo numero di prodotti attesi (subarea bibliometrica)

Tabella 1.10 - Composizione del GEV 11

Tabella 1.11 - Prodotti gestiti da ciascun componente del GEV 11

Tabella 1.12 - Riunioni GEV 11

Tabella 2.1 - Numero di prodotti conferiti, attesi e mancanti per SSD (sub-area non bibliometrica)

Tabella 2.2 - Numero di prodotti conferiti, attesi e mancanti per SSD (sub-area bibliometrica)

Tabella 2.3 - Revisori utilizzati per l'area 11, per numero di revisioni effettuate

Tabella 2.4 - Percentuale revisioni inevase o rifiutate in area 11 (calcolata sul totale di quelle richieste)

Tabella 2.5 - Revisioni richieste, effettuate ed inevase da revisori italiani e stranieri per SSD, area 11



Tabella 2.6 - Totale revisioni richieste, effettuate ed inevase per SSD, area 11

Tabella 2.7 - Revisori con bassa varianza normalizzata (calcolati su quanti hanno effettuato almeno 8 revisioni)

Tabella 2.8 - Distribuzione valutazioni per fascia di docenza, Discipline Demoeetnoantropologiche (percentuali calcolate sui prodotti effettivamente valutati)

Tabella 2.9 - Distribuzione valutazioni per fascia di docenza, Geografia (percentuali calcolate sui prodotti effettivamente valutati)

Tabella 2.10 - Distribuzione valutazioni per fascia di docenza, Scienze del Libro e del Documento (percentuali calcolate sui prodotti effettivamente valutati)

Tabella 2.11 - Distribuzione valutazioni per fascia di docenza, Storia (percentuali calcolate sui prodotti effettivamente valutati)

Tabella 2.12 - Distribuzione valutazioni per fascia di docenza, Filosofia (percentuali calcolate sui prodotti effettivamente valutati)

Tabella 2.13 - Distribuzione valutazioni per fascia di docenza, Pedagogia (percentuali calcolate sui prodotti effettivamente valutati)

Tabella 2.14 - Distribuzione valutazioni per fascia di docenza, sub-area bibliometrica (percentuali calcolate sui prodotti effettivamente valutati)

Tabella 2.15 - Prodotti con revisioni fortemente discordanti (E vs. L), sub-area non bibliometrica. Percentuali calcolate sui prodotti effettivamente valutati, per SSD di valutazione del prodotto

Tabella 2.16 - Prodotti con revisioni fortemente discordanti (E vs. L), sub-area bibliometrica. Percentuali calcolate sui prodotti effettivamente valutati, per SSD di valutazione del prodotto

Tabella 2.17 - Numeri e percentuali di valutazioni concordanti e minimamente divergenti, sub-area non bibliometrica. Percentuali calcolate sui prodotti effettivamente valutati, per SSD di valutazione del prodotto



Tabella 2.18 - Numeri e percentuali di valutazioni concordanti e minimamente divergenti, sub-area bibliometrica. Percentuali calcolate sui prodotti effettivamente valutati, per SSD di valutazione del prodotto

Tabella 2.19: Percentuale di prodotti di ricerca valutati attraverso la peer review, la bibliometria e con entrambi i metodi di valutazione (sub-area bibliometrica)

Tabella 2.20: Punteggi ottenuti e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per subGEV, sub-area non bibliometrica.

Tabella 2.21: Punteggi ottenuti e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per subGEV, sub-area bibliometrica.

Tabella 2.22: Punteggi ottenuti e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per SSD, sub-area non bibliometrica.

Tabella 2.23: Punteggi ottenuti e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per SSD, sub-area bibliometrica

Tabella 2.24: Punteggi ottenuti e distribuzione dei prodotti conferiti non penalizzati nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per tipologia di pubblicazione, sub-area non bibliometrica

Tabella 2.25: Punteggi ottenuti e distribuzione dei prodotti conferiti non penalizzati nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per tipologia di pubblicazione, sub-area bibliometrica

Tabella 2.26 - Punteggi ottenuti e distribuzione dei prodotti conferiti non penalizzati nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per tipologia di prodotto (indicizzato o no in WoS/Scopus), sub-area bibliometrica

Tabella 2.27 - Distribuzione delle valutazioni per fasce di classificazione delle riviste: percentuali di E-B-A-L rispetto al totale degli articoli per fascia di classificazione

Tabella 2.28 - Distribuzione delle valutazioni su riviste nazionali e internazionali: percentuali di E-B-A-L rispetto al totale degli articoli

Tabella 2.29 - Distribuzione delle valutazioni sulle riviste internazionali: percentuali di E-B-A-L rispetto al totale degli articoli



Tabella 2.30 - Distribuzione degli articoli sulle riviste internazionali rispetto al totale - percentuali sul totale di articoli E-B-A-L

Tabella 2.31 - Distribuzione delle valutazioni per fasce di classificazione (rivedute) delle riviste: percentuali di E-B-A-L sul totale degli articoli

Tabella 2.32 - Distribuzione delle valutazioni per fasce di classificazione delle riviste: percentuali per fascia di classificazione sull'insieme degli articoli E-B-A-L

Tabella 2.33 - Distribuzione delle valutazioni sulle riviste di fascia A, Discipline Demoetnoantropologiche

Tabella 2.34 - Distribuzione delle valutazioni sulle riviste di fascia A, Filosofia

Tabella 2.35 - Distribuzione delle valutazioni sulle riviste di fascia A, Geografia

Tabella 2.36 - Distribuzione delle valutazioni sulle riviste di fascia A, Pedagogia

Tabella 2.37 - Distribuzione delle valutazioni sulle riviste di fascia A, Scienze del Libro e del Documento

Tabella 2.38 - Distribuzione delle valutazioni sulle riviste di fascia A, Storia

Tabella 3.1 - Distribuzione dei prodotti Eccellenti tra le strutture, Discipline Demoetnoantropologiche

Tabella 3.2 - Distribuzione dei prodotti Eccellenti tra le strutture, Filosofia Teoretica

Tabella 3.3 - Distribuzione dei prodotti Eccellenti tra le strutture, Logica e Filosofia della Scienza

Tabella 3.4 - Distribuzione dei prodotti Eccellenti tra le strutture, Filosofia Morale

Tabella 3.5 - Distribuzione dei prodotti Eccellenti tra le strutture, Estetica

Tabella 3.6 - Distribuzione dei prodotti Eccellenti tra le strutture, Filosofia e Teoria dei Linguaggi

Tabella 3.7 - Distribuzione dei prodotti Eccellenti tra le strutture, Storia della Filosofia

Tabella 3.8 - Distribuzione dei prodotti Eccellenti tra le strutture, Storia della Filosofia antica



Tabella 3.9 - Distribuzione dei prodotti Eccellenti tra le strutture, Storia della Filosofia medievale

Tabella 3.10 - Distribuzione dei prodotti Eccellenti tra le strutture, Storia della Scienza e delle Tecniche

Tabella 3.11 - Distribuzione dei prodotti Eccellenti tra le strutture, Geografia

Tabella 3.12 - Distribuzione dei prodotti Eccellenti tra le strutture, Geografia Economico-Politica

Tabella 3.13 - Distribuzione dei prodotti Eccellenti tra le strutture, Pedagogia Generale e Sociale

Tabella 3.14 - Distribuzione dei prodotti Eccellenti tra le strutture, Storia della Pedagogia

Tabella 3.15 - Distribuzione dei prodotti Eccellenti tra le strutture, Didattica e Pedagogia Speciale

Tabella 3.16 - Distribuzione dei prodotti Eccellenti tra le strutture, Pedagogia Sperimentale

Tabella 3.17 - Distribuzione dei prodotti Eccellenti tra le strutture, Archivistica-Bibliografia-Biblioteconomia

Tabella 3.18 - Distribuzione dei prodotti Eccellenti tra le strutture, Paleografia

Tabella 3.19 - Distribuzione dei prodotti Eccellenti tra le strutture, Storia Medievale

Tabella 3.20 - Distribuzione dei prodotti Eccellenti tra le strutture, Storia Moderna

Tabella 3.21 - Distribuzione dei prodotti Eccellenti tra le strutture, Storia dell'Europa orientale

Tabella 3.22 - Distribuzione dei prodotti Eccellenti tra le strutture, Storia Contemporanea

Tabella 3.23 - Distribuzione dei prodotti Eccellenti tra le strutture, Storia delle Religioni

Tabella 3.24 - Distribuzione dei prodotti Eccellenti tra le strutture, Storia del Cristianesimo e delle Chiese

Tabella 3.25 - Distribuzione dei prodotti Eccellenti tra le strutture, Psicologia Generale

Tabella 3.26 - Distribuzione dei prodotti Eccellenti tra le strutture, Psicobiologia e Psicologia Fisiologica



Tabella 3.27 - Distribuzione dei prodotti Eccellenti tra le strutture, Psicometria

Tabella 3.28 - Distribuzione dei prodotti Eccellenti tra le strutture, Psicologia dello Sviluppo e Psicologia dell'Educazione

Tabella 3.29 - Distribuzione dei prodotti Eccellenti tra le strutture, Psicologia Sociale

Tabella 3.30 - Distribuzione dei prodotti Eccellenti tra le strutture, Psicologia del Lavoro e delle Organizzazioni

Tabella 3.31 - Distribuzione dei prodotti Eccellenti tra le strutture, Psicologia Dinamica

Tabella 3.32 - Distribuzione dei prodotti Eccellenti tra le strutture, Psicologia Clinica

Tabella 3.34 - Distribuzione dei prodotti Eccellenti tra le strutture, Metodi e Didattiche delle Attività Sportive

Tabella 3.35 - Numero dei soggetti valutati attivi, non attivi, parzialmente attivi e parzialmente inattivi per struttura (sub-area non bibliometrica)

Tabella 3.36 - Numero soggetti valutati attivi, non attivi, parzialmente attivi e parzialmente inattivi per struttura (sub-area bibliometrica)

Tabella 3.37: Graduatorie delle Università divise per dimensione in base al numero di prodotti attesi dall'Area 11 (sub-area non bibliometrica)

Tabella 3.38: Graduatorie delle Università divise per dimensione in base al numero di prodotti attesi dall'Area 11 (sub-area bibliometrica)

Tabella 3.39: Graduatoria degli Enti di ricerca dell'Area 11 (sub-area non bibliometrica) ordinata per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.40: Graduatoria degli Enti di ricerca dell'Area 11 (sub-area bibliometrica) ordinata per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.41 - Graduatoria di Università per Discipline Demoeconomiche, ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)



Tabella 3.42 - Graduatoria di Università per Scienze del Libro e del Documento, ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.43 - Graduatoria di Università per Filosofia, divise per dimensione in base al numero di prodotti attesi: G=grandi ($n \geq 100$), M=medie ($99 \geq n \geq 30$) e P=piccole ($n < 30$), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.44 - Graduatoria di Università per Geografia, ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.45 - Graduatoria di Università per Pedagogia (settori non bibliometrici), divise per dimensione in base al numero di prodotti attesi: G=grandi ($n \geq 100$), M=medie ($99 \geq n \geq 30$) e P=piccole ($n < 30$) e ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.46 - Graduatoria di strutture per Storia, divise per dimensione in base al numero di prodotti attesi: G=grandi ($n \geq 100$), M=medie ($99 \geq n \geq 30$) e P=piccole ($n < 30$), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.47 - Graduatoria di strutture per Psicologia, divise per dimensione in base al numero di prodotti attesi: G=grandi ($n \geq 100$), M=medie ($99 \geq n \geq 30$) e P=piccole ($n < 30$), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.48 - Graduatoria di strutture per Metodi e Didattiche delle Attività Motorie e Sportive, ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.49 - Graduatoria di strutture per Filosofia Teoretica (M-FIL/01), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)



Tabella 3.50 - Graduatoria di strutture per Logica e Filosofia della Scienza (M-FIL/02), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.51 - Graduatoria di strutture per Filosofia Morale (M-FIL/03), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.52 - Graduatoria di strutture per Estetica (M-FIL/04), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.53 - Graduatoria di strutture per Filosofia e Teoria dei Linguaggi (M-FIL/05), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.54 - Graduatoria di strutture per Storia della Filosofia (M-FIL/06), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.55 - Graduatoria di strutture per Geografia (M-GGR/01), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.56 - Graduatoria di strutture per Geografia Economico Politica (M-GGR/02), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.57 - Graduatoria di strutture per Pedagogia Generale e Sociale (M-PED/01), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.58 - Graduatoria di strutture per Storia della Pedagogia (M-PED/02), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)



Tabella 3.59 - Graduatoria di strutture per Didattica e Pedagogia Speciale (M-PED/03), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.60 - Graduatoria di strutture per Pedagogia Sperimentale (M-PED/04), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.61 - Graduatoria di strutture per Storia Medievale (M-STO/01), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.62 - Graduatoria di strutture per Storia Moderna (M-STO/02), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.63 - Graduatoria di strutture per Storia dell'Europa orientale (M-STO/03), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.64 - Graduatoria di strutture per Storia Contemporanea (M-STO/04), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.65 - Graduatoria di strutture per Storia della Scienza e delle Tecniche (M-STO/05), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.66 - Graduatoria di strutture per Storia del Cristianesimo e delle Chiese (M-STO/07), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.67 - Graduatoria di strutture per Archivistica, Bibliografia e Biblioteconomia (M-STO/08), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)



Tabella 3.68 - Graduatoria di strutture per Paleografia (M-STO/09), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.69 - Graduatoria di strutture per Metodi e Didattiche delle Attività Motorie (M-EDF/01), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.70 - Graduatoria di strutture per Metodi e Didattiche delle Attività Sportive (M-EDF/02), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.71 - Graduatoria di strutture per Psicologia Generale (M-PSI/01), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.72 - Graduatoria di strutture per Psicobiologia e Psicologia Fisiologica (M-PSI/02), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.73 - Graduatoria di strutture per Psicometria (M-PSI/03), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.74 - Graduatoria di strutture per Psicologia dello Sviluppo e dell'Educazione (M-PSI/04), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.75 - Graduatoria di strutture per Psicologia Sociale (M-PSI/05), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.76 - Graduatoria di strutture per Psicologia del Lavoro e delle Organizzazioni (M-PSI/06), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)



Tabella 3.77 - Graduatoria di strutture per Psicologia Dinamica (M-PSI/07), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.78 - Graduatoria di strutture per Psicologia Clinica (M-PSI/08), ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 3.79: Graduatorie delle Università divise per dimensione in base al numero di prodotti attesi dall'Area 11 (sub-area non bibliometrica): G=grandi ($n \geq 300$), M=medie ($299 \geq n \geq 100$) e P=piccole ($n < 100$) e ordinate per IRAS1

Tabella 3.80: Graduatorie delle Università divise per dimensione in base al numero di prodotti attesi dall'Area 11 (sub-area bibliometrica): G=grandi ($n \geq 100$), M=medie ($99 \geq n \geq 30$) e P=piccole ($n < 30$) e ordinate per IRAS1

Tabella 3.81: Graduatoria degli Enti di ricerca dell'Area 11 (sub-area bibliometrica) ordinata per IRAS1

Tabella 4.1 - Graduatorie dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), divisi tra Grandi (G, con più di 100 prodotti), Medi (M, da 30 a 99 prodotti) e Piccoli (P, sotto i 29 prodotti), sub-area non bibliometrica

Tabella 4.2 - Graduatorie dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), divisi tra Grandi (G, con più di 100 prodotti), Medi (M, da 30 a 99 prodotti) e Piccoli (P, sotto i 29 prodotti), sub-area bibliometrica

Tabella 4.3 - Graduatorie dei dipartimenti pre L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), divisi tra Grandi (G, con più di 100 prodotti), Medi (M, da 30 a 99 prodotti) e Piccoli (P, sotto i 29 prodotti), sub-area non bibliometrica

Tabella 4.4- Graduatorie dei dipartimenti pre L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), divisi tra Grandi (G, con più di 100 prodotti), Medi (M, da 30 a 99 prodotti) e Piccoli (P, sotto i 29 prodotti), sub-area bibliometrica



Tabella 4.5 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Discipline Demoetnoantropologiche

Tabella 4.6 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), divisi tra divisi tra Grandi (G, con almeno 30 prodotti) e Piccoli (P, sotto i 29 prodotti), per Filosofia

Tabella 4.7 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Geografia

Tabella 4.8 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), divisi tra divisi tra Grandi (G, con almeno 30 prodotti) e Piccoli (P, sotto i 29 prodotti), per Pedagogia (settori non bibliometrici)

Tabella 4.9 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Scienze del Libro e del Documento

Tabella 4.10 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), divisi tra divisi tra Grandi (G, con almeno 30 prodotti) e Piccoli (P, sotto i 29 prodotti), per Storia

Tabella 4.11 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Metodi e Didattiche delle Attività Motorie e Sportive

Tabella 4.12 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), divisi tra divisi tra Grandi (G, con almeno 30 prodotti) e Piccoli (P, sotto i 29 prodotti), per Psicologia



Tabella 4.13 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Filosofia Teoretica (M-FIL/01)

Tabella 4.14 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Logica e Filosofia della Scienza (M-FIL/02)

Tabella 4.15 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Filosofia Morale (M-FIL/03)

Tabella 4.16 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Estetica (M-FIL/04)

Tabella 4.17 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Filosofia e Teoria dei Linguaggi (M-FIL/05)

Tabella 4.18 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Storia della Filosofia (M-FIL/06)

Tabella 4.19 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Storia della Scienza e delle Tecniche (M-STO/05)

Tabella 4.20 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Geografia (M-GGR/01)

Tabella 4.21 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Geografia Economico Politica (M-GGR/02)



Tabella 4.22 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Pedagogia Generale e Sociale (M-PED/01)

Tabella 4.23 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Storia della Pedagogia (M-PED/02)

Tabella 4.24 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Pedagogia e Didattica Speciale (M-PED/03)

Tabella 4.25 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Pedagogia Sperimentale (M-PED/04)

Tabella 4.26 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Storia Medievale (M-STO/01).

Tabella 4.27 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Storia Moderna (M-STO/02)

Tabella 4.28 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Storia Contemporanea (M-STO/04)

Tabella 4.29 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Storia del Cristianesimo e delle Chiese (M-STO/07)

Tabella 4.30 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Archivistica Bibliografia e Biblioteconomia (M-STO/08)



Tabella 4.31 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Paleografia (M-STO/09)

Tabella 4.32 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Metodi e Didattiche delle Attività Motorie (M-EDF/01)

Tabella 4.33 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Metodi e Didattiche delle Attività Sportive (M-EDF/02)

Tabella 4.34 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Psicologia Generale (M-PSI/01)

Tabella 4.35 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Psicobiologia e Psicologia Fisiologica (M-PSI/02)

Tabella 4.36 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Psicometria (M-PSI/03)

Tabella 4.37 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Psicologia dello Sviluppo e Psicologia dell'Educazione (M-PSI/04)

Tabella 4.38 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Psicologia Sociale (M-PSI/05)

Tabella 4.39 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Psicologia del Lavoro e delle Organizzazioni (M-PSI/06)



Tabella 4.40- Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Psicologia Dinamica (M-PSI/07)

Tabella 4.41 - Graduatoria dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Psicologia Clinica (M-PSI/08)

Tabella 4.42 - Graduatorie dei dipartimenti post L.240 ordinati per voto medio ($I=v/n$) all'interno delle strutture (elencate in ordine alfabetico) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), sub-area non bibliometrica

Tabella 4.43 - Graduatorie dei dipartimenti post L.240 ordinati per IRD1 all'interno delle strutture (elencate in ordine alfabetico), sub-area non bibliometrica

Tabella 4.44 - Graduatorie dei dipartimenti post L.240 ordinati per voto medio ($I=v/n$) all'interno delle strutture (elencate in ordine alfabetico) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), sub-area bibliometrica

Tabella 4.45 - Graduatorie dei dipartimenti post L.240 ordinati per IRD1 all'interno delle strutture (elencate in ordine alfabetico), sub-area bibliometrica

Tabella 4.46 - Graduatorie dei dipartimenti pre L.240 ordinati per voto medio ($I=v/n$) all'interno delle strutture (elencate in ordine alfabetico) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), sub-area non bibliometrica.

Tabella 4.47 - Graduatorie dei dipartimenti pre L.240 ordinati per IRD1 all'interno delle strutture (elencate in ordine alfabetico), sub-area non bibliometrica.

Tabella 4.48 - Graduatorie dei dipartimenti pre L.240 ordinati per voto medio ($I=v/n$) all'interno delle strutture (elencate in ordine alfabetico) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), sub-area bibliometrica

Tabella 4.49 - Graduatorie dei dipartimenti pre L.240 ordinati per IRD1 all'interno delle strutture (elencate in ordine alfabetico), sub-area bibliometrica.

Tabella 4.50 - Graduatoria delle sottostrutture degli enti di ricerca, ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), sub-area non bibliometrica

Tabella 4.51 - Graduatoria delle sottostrutture degli enti di ricerca, ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-), sub-area bibliometrica

Tabella 4.52: Graduatorie delle sottostrutture degli enti di ricerca, ordinate per voto medio ($I=v/n$) all'interno delle strutture (elencate in ordine alfabetico) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-)

Tabella 4.53 - Graduatoria delle sottostrutture degli Enti ricerca, ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Psicologia.

Tabella 4.54 - Graduatoria delle sottostrutture degli Enti ricerca, ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per Psicobiologia e Psicologia fisiologica (M-PSI/02)

Tabella 5.1 - Graduatoria delle strutture per voto medio (VM) dei soggetti valutati (SV) e distribuzione in classi di voto medio, Area 11 (sub-area non bibliometrica)

Tabella 5.2 - Graduatoria delle strutture per voto medio (VM) dei soggetti valutati (SV) e distribuzione in classi di voto medio, Area 11 (sub-area bibliometrica)

Tabella 5.3 - Graduatoria dei dipartimenti ex L.240 per voto medio (VM) dei soggetti valutati (SV) e distribuzione in classi di voto medio, Area 11 (sub-area non bibliometrica)

Tabella 5.4 - Graduatoria dei dipartimenti ex L.240 per voto medio (VM) dei soggetti valutati (SV) e distribuzione in classi di voto medio, Area 11 (sub-area bibliometrica)

Tabella 5.5 - Punteggio medio dei soggetti valutati per subGEV e anno di assunzione in ruolo, area 11 (sub-area non bibliometrica)

Tabella 5.6 - Punteggio medio dei soggetti valutati per subGEV e anno di assunzione in ruolo, area 11 (sub-area bibliometrica)

Tabella 5.7 - Distribuzione geografica dei prodotti E-B-A-L, sub-area non bibliometrica

Tabella 5.8 - Distribuzione geografica dei prodotti E per SSD (sub-area non bibliometrica)

Tabella 5.9 - Distribuzione geografica dei prodotti B per SSD (sub-area non bibliometrica)



Tabella 5.10 - Distribuzione geografica dei prodotti A per SSD (sub-area non bibliometrica)

Tabella 5.11 - Distribuzione geografica dei prodotti L per SSD (sub-area non bibliometrica)

Tabella 5.12 - Distribuzione geografica dei prodotti E-B-A-L, sub-area bibliometrica

Tabella 5.13 - Distribuzione geografica dei prodotti E per SSD (sub-area bibliometrica)

Tabella 5.14 - Distribuzione geografica dei prodotti B per SSD (sub-area bibliometrica)

Tabella 5.15 - Distribuzione geografica dei prodotti A per SSD (sub-area bibliometrica)

Tabella 5.16 - Distribuzione geografica dei prodotti L per SSD (sub-area bibliometrica)

National Agency for the Evaluation of
Universities and Research Institutes



Agenzia Nazionale di Valutazione del
sistema Universitario e della Ricerca

Evaluation of Research Quality



Valutazione Qualità della Ricerca



1 Introduzione

L'Area 11 è una realtà composta, che riunisce 38 diversi settori scientifico-disciplinari (SSD) appartenenti ad una varietà di macro-settori scientifici: Antropologia, Filosofia, Geografia, Pedagogia, Psicologia, Storia (cfr. Tabella 1.1). In altri sistemi nazionali, come quello inglese, e in Europa, come dimostra la strutturazione dell'European Research Council, tali macro-settori godono di piena autonomia, essendo muniti di tradizioni e culture proprie, riflessi in una notevole variabilità di tipi di pubblicazione, che vanno dagli articoli su riviste internazionali presenti nelle maggiori banche dati, a quelli su riviste non presenti in tali banche dati, alle monografie e ai capitoli di libro pubblicati in Italia o all'estero, alle edizioni di fonti (cfr. in proposito le Tabelle 1.2-1.5).

Al Gruppo Esperti Valutatori (GEV 11) è stato chiesto di valutare 13.152 lavori di 4.749 soggetti valutati appartenenti all'Area 11 o comunque ad essa riconducibili attraverso le scelte da loro fatte al momento dell'invio dei prodotti selezionati per la VQR, come nel caso del personale degli enti di ricerca. Nel far ciò il GEV 11 ha utilizzato la griglia di classificazione fissata dal bando ministeriale, secondo la quale ogni lavoro andava collocato in una delle quattro classi di merito previste (E-B-A-L, Eccellenti, Buoni, Accettabili, Limitati) attraverso l'analisi bibliometrica e/o la *peer review*. Esso ha agito tenendo ferma la finalità esclusiva della VQR, che consiste nella valutazione delle strutture e non dei singoli.

Le tradizioni culturali, la mancanza di banche dati rappresentative e la tipologia della produzione della maggioranza dei macro-settori presenti nell'Area 11 hanno reso subito evidente la necessità di ricorrere in modo prevalente alla *peer review*. L'eccezione è stata rappresentata dal macro-settore della Psicologia e dai due Settori Scientifico-Disciplinari (SSD) di *Metodi e didattiche delle attività motorie e sportive* (M-EDF 1 e 2), che hanno ritenuto la bibliometria più adatta alla tipologia della loro produzione scientifica.

Allo stesso tempo, la componente non bibliometrica del GEV 11 ha avuto subito chiara coscienza di alcuni dei limiti e dei problemi sollevati dalla *peer review*: la dimensione stessa della VQR, che implicava la doppia valutazione di oltre 10.000 lavori, e quindi la gestione di oltre 20.000 pareri, sembrava infatti ridurre l'affidabilità della *peer review* a causa tanto dell'altissimo numero di *referee* che era necessario coinvolgere, quanto dell'enorme domanda di tempo e attenzione che l'elevato numero dei prodotti da valutare poneva ai *referee* coinvolti. Nonostante la cura posta nella selezione dei *referee* e nell'assegnazione dei prodotti, la *peer review*—metodo per eccellenza soggettivo—non può inoltre escludere del tutto i conflitti di



interesse, per quanto si affronti il problema e si approntino rimedi, né soprattutto assicurare, per lo stesso motivo, una valutazione uniforme tra i diversi lavori, settori disciplinari, temi di ricerca, ecc. Si intravedeva infine il problema posto dal dover ricondurre a sintesi gli eventuali, e probabili, giudizi difformi dei revisori, soprattutto nei casi in cui i loro pareri fossero risultati molto discordi. A questi problemi altri se ne sono poi via via aggiunti nel corso del processo di valutazione, sollevando un dibattito di cui si darà conto nel Rapporto.

Il GEV 11 ha affrontato il lavoro convinto che sia interesse preminente del paese far sì che tutte le aree del sapere vi siano rappresentate e coltivate nel miglior modo possibile. La VQR insomma non può che proporsi il fine di stimolare il miglioramento della qualità della produzione scientifica in tutte le aree e le discipline in cui il sapere è articolato, individuando e premiando le loro parti più vive, in modo da dar loro fiducia e risorse.

Anche laddove fosse possibile—e ciò sembra contestabile, una comparazione tra aree e discipline tesa a premiarne alcune piuttosto che altre sarebbe quindi sbagliata, perché danneggerebbe la ricchezza e la complessità di una cultura nazionale che può e deve aspirare a integrarsi con la migliore cultura internazionale in tutti i campi del sapere, e non può correre il rischio di perdere dei pezzi. Solo l'autorità politica democratica potrebbe di volta in volta scegliere su quali settori puntare, ricordando tuttavia che spesso, almeno nel campo della ricerca pura, la pianificazione ha prodotto risultati deludenti, a ragione dell'imprevedibilità della direzione e degli esiti di tale ricerca.

Nel suo complesso, l'esperienza della VQR è stata positiva anche per il vivace dibattito che ha generato, permettendo alla nostra comunità scientifica di affrontare in maniera diretta e organica questioni prima almeno in parte sottovalutate: la valutazione della qualità del lavoro fatto anche alla luce del suo riconoscimento nazionale e internazionale; la presa di coscienza dei limiti intrinseci ad un isolamento certo mai assoluto ma talvolta eccessivo; una più acuta percezione dell'esistenza di gerarchie basate sul merito che costituiscono un elemento imprescindibile di una attività, come la ricerca, basata sul desiderio di far progredire il sapere nella maniera più significativa e originale possibile, sono state altrettanto conquiste di un settore—quello della ricerca italiana—che è stato l'unico del paese non solo ad accettare di sottoporsi a un grande processo di valutazione, ma anche a partecipare attivamente ad esso.

Questa partecipazione attiva e intelligente della comunità scientifica nel suo complesso è stata forse la caratteristica più stimolante e promettente della VQR in Area 11. La partecipazione e il sostegno dell'ANVUR erano scontati, ma lo erano meno la loro qualità, che è stata invece



molto alta. Prevedibile era anche la collaborazione del GEV, ma anche in questo caso era difficile immaginare che il dibattito interno sarebbe stato così vivace e al tempo stesso costruttivo. Di tutto ciò vanno ringraziati tutti i membri del GEV, e in particolare i quattro coordinatori di sub-GEV (Francesco Benigno per Antropologia, Geografia e Storia; Diego Marconi per la Filosofia; Giuseppe Bertagna per la Pedagogia e Stefano Puglisi-Allegra per la Psicologia) e l'assistente del GEV, Dr. Antonio Ferrara.

Molto proficua è stata la collaborazione con le Società scientifiche e le Consulte dell'Area, che ha permesso di svolgere il lavoro rispettando le caratteristiche e le tradizioni di ogni disciplina, senza per questo rinunciare a un rigore e a una serietà, intesi con equilibrio, che le stesse Società e le Consulte hanno subito capito essere nell'interesse comune. Ciò vale anche nell'unico caso—quello della Filosofia Teoretica—in cui non è stato purtroppo possibile raggiungere un accordo per insuperabili divergenze sulla possibilità e opportunità di una classificazione di merito delle riviste, senza comunque che questo impedisse un dialogo sempre interessante. E' quindi con grande piacere che ne ringraziamo i Presidenti, che hanno permesso alla VQR di svolgersi col consenso e l'appoggio delle singole comunità scientifiche, un consenso e un appoggio dimostrati dai circa 1.440 *referee* coinvolti (dei quali circa 1.330 sono stati nominati dal GEV-11 stesso) e soprattutto dall'estrema limitatezza dei rifiuti a svolgere tale attività. Quelli motivati dalle critiche alla valutazione si sono contati sulle dita di una mano, e anche quelli legati agli impegni e alla scarsità del tempo a disposizione sono stati molto contenuti, specie tra gli studiosi italiani, a conferma della percezione dell'importanza della VQR per il nostro futuro. Almeno in Area 11 la VQR ha quindi ricevuto una forte legittimazione dalla comunità scientifica, che ha compreso che si trattava di un'importante occasione di miglioramento e autodisciplina, di cui attende ora le ricadute positive.

Anche gli editori hanno giocato un ruolo essenziale, dimostrando grande disponibilità nella messa a disposizione—spesso finanziariamente gravosa—delle migliaia di PDF indispensabili al successo della VQR, PDF che poi il CINECA è riuscito, sia pure con qualche difficoltà, a rendere visibili a migliaia di *referee*. Chi scrive nutreva dubbi sul successo di questa operazione, e non può quindi che ringraziare chi lo ha reso possibile—vale a dire editori e CINECA—malgrado quelle difficoltà, forse inevitabili data la scala dell'impresa. Si può a proposito notare che la VQR ha giocato un ruolo cruciale nell'accelerare la messa *on line* della ricerca italiana, un'operazione poi continuata con le Abilitazioni scientifiche nazionali. Questa messa *on line* ha raggiunto in poco tempo dimensioni prima impensabili, con conseguenze fondamentali nel campo della trasparenza delle procedure accademiche, e rappresenta di per sé uno dei successi della VQR.



Grazie a questi attori è stato possibile portare a termine 23.062 referaggi (di cui 21.358 effettuati da revisori esterni al GEV) e 2.146 valutazioni bibliometriche (quelle di Psicologia e dei settori M-EDF) nei tempi stabiliti, nonostante i circa tre mesi di slittamento nell'avvio del processo di valutazione dipesi dalla richiesta di Università e enti di ricerca di avere più tempo a disposizione per la selezione e l'invio dei prodotti da valutare, e dalle procedure della Abilitazione scientifica nazionale che hanno assorbito per qualche settimana la maggior parte delle risorse umane e materiali del CINECA, e in parte anche dell'ANVUR. E' stato perciò possibile trattare i prodotti ricevuti solo a partire dal settembre 2012, e a fine febbraio 2013 il processo era sostanzialmente concluso.

Come ha scritto un membro del GEV, la VQR è stata quindi davvero uno “sforzo eccezionale... sul piano scientifico, statistico, antropologico e sociologico”, e siamo certi che l'analisi dei suoi risultati, di cui questo rapporto è solo la prima tappa, si rivelerà utile per il paese e la sua comunità scientifica.

Questa analisi ci dice che se è senz'altro possibile parlare di successo della VQR, quest'ultima ha fatto emergere problemi metodologici che non vanno sottovalutati. *Il principale, segnalato in Area 11 dalla diversità tra l'esito della valutazione in Psicologia e nei settori M-EDF, e quello di tutti gli altri SSD, è la differenza strutturale tra i risultati della peer review e quelli della bibliometria. La prima si è infatti mostrata, con poche eccezioni facilmente spiegabili, meno generosa della seconda per quanto riguarda i prodotti eccellenti, con percentuali che in media si aggirano intorno al 15 % a fronte di una media di circa il 35% nella bibliometria, e allo stesso tempo meno propensa ad assegnare i prodotti valutati alla classe inferiore (Limitati), con percentuali medie al di sotto del 19% contro il 27% della bibliometria.* Il fenomeno, impreveduto, era in fin dei conti prevedibile una volta che si rifletta sulle sue cause: la *peer review* è infatti un sistema soggettivo, ed è molto difficile che delle persone (vale a dire i *referee*) assegnino—per motivi che sono anche psicologici—alla fasce estreme, e in particolare a quella dell'eccellenza più del 10-15% dei prodotti, che è poi il dato medio degli SSD non bibliometrici, eccezion fatta per quelli più piccoli, di cui si dirà tra breve. A differenza della bibliometria, che permette agli autori di scegliere i prodotti da valutare in base a criteri precisi, che ne garantiscono il risultato (per esempio la classificazione in E), la *peer review* ha inoltre un alto margine di incertezza, dovuto appunto alla soggettività dei valutatori, che rende impossibile agli autori essere sicuri a priori dei risultati della loro selezione dei prodotti da presentare, e abbassa quindi la percentuale di E conseguite.



I risultati, strutturalmente diversi, di una valutazione peer rispetto a una valutazione bibliometrica, NON possono e NON debbono quindi essere confrontati, perché se ne trarrebbero indicazioni errate. A causa di questa incomparabilità strutturale (su cui torneremo in 5. *Analisi dei risultati*) l'Area 11 presenta i risultati della VQR come se fosse costituita da due sub-aree distinte, di cui uno ha fatto prevalentemente ricorso alle valutazioni bibliometriche e l'altra invece ha mandato in *peer review* tutti i prodotti che le sono stati sottoposti. Solo l'indicatore IRAS è stato calcolato a livello di area anziché di sub-aree *peer* e bibliometrica, poiché esso mette insieme qualità e quantità, ed è quindi funzionale alla ripartizione del FFO, che riguarda le sole strutture.

Un problema diverso, ma anch'esso di ordine generale, è rappresentato dalla notevole variabilità emersa nei risultati della valutazione *peer* tra SSD anche appartenenti allo stesso macro-settore e persino allo stesso Settore Concorsuale (SC). Ciò a riprova della soggettività di uno strumento che esalta, proprio a causa di questa sua natura soggettiva, la diversa cultura e le diverse tradizioni, anche in materia di valutazione (accolta con maggiore o minore adesione o diffidenza), proprie di ciascun SSD.

Un problema simile è posto dai SSD più piccoli, che hanno ottenuto risultati sistematicamente e notevolmente migliori di quelli degli SSD più grandi. Proprio la sistematicità di questa differenza rende difficile capire se i migliori (e spesso i molto migliori) risultati ottenuti dai piccoli SSD dipendano o meno da una loro reale migliore qualità rispetto ai grandi SSD, come potrebbe essere in qualche caso possibile (anche qui vedi 5. *Analisi dei risultati*).

La separazione tra SSD bibliometrici e non in Area 11 segnala infine a sua volta un ultimo problema di carattere generale che la VQR ha permesso di mettere meglio a fuoco. Si tratta dell'inadeguatezza, almeno ai fini della valutazione, delle aree CUN, spesso frutto di stratificazioni storiche, esigenze didattiche e talvolta anche corporative, e della necessità di trovare comunque una collocazione a SSD, e persino a intere discipline, che era difficile situare in un sistema con una base così ristretta (ricordiamo che a fronte delle 14 aree italiane stanno per esempio le 36 della valutazione britannica, dove antropologia, pedagogia e geografia godono di autonomia nell'ambito, peraltro, di un *Main Panel* diverso da quello cui afferiscono storia e filosofia).

Problemi particolari sono invece emersi relativamente alla scheda di valutazione preparata per i *referee*. Molti revisori e membri del GEV hanno concentrato la loro attenzione su due punti, segnalando la necessità di trovare il modo di dare in futuro maggior rilievo alla



consistenza dei prodotti da valutare (l'ovvio caso è quello delle monografie rispetto agli articoli) e di arrivare a una migliore definizione dell'internazionalizzazione, che è stata spesso interpretata in modo difforme, rivelandosi una delle fonti più importanti di disparità nella valutazione di uno stesso prodotto ad opera di *referee* diversi.

Nelle *Conclusioni* torneremo sulle lezioni che si possono trarre da questi problemi, di cui è opportuno però qui anticipare, usando le parole di un membro del GEV, quella che sembra la principale, vale a dire la necessità di assumere un atteggiamento critico rispetto “a un sistema valutativo, quello della *peer*” che pure “sembrava, sino a un anno fa, impossibile da sostituire o da non difendere... Qualunque garante serio non può non essersi accorto di quanto le variabili soggettive—nel bene e nel male—abbiano inciso nella valutazione”. Critico non vuole ovviamente dire liquidatorio, ma piuttosto teso al miglioramento di questo sistema anche attraverso una sua integrazione: si tratta insomma di prendere coscienza dei limiti di una valutazione puramente *peer* e della necessità di renderla più sofisticata, pensando a “modi valutativi che, pur non sostituendola, elaborino criteri ‘quantitativi’ in grado almeno di essere comparati a quelli qualitativi e soggettivi”, correggendo così, “la varietà psicologica e accademica dei valutatori”, quelle “arbitrarie discrepanze” di cui ha parlato un altro EV riflettendo sulla valutazione nel suo SSD.

Questa necessità sembra confermata, nelle osservazioni di altri membri GEV, dalle più convincenti distribuzioni statistiche delle valutazioni dei due SSD bibliometrici (M-EDF/01 e 02) associati alla Pedagogia rispetto alle valutazioni tramite *peer review* degli SSD propriamente pedagogici, e da quanto è accaduto in quella parte della Psicologia (ma è probabile che ciò sia avvenuto in tutti i macro-settori dell'Area 11) che ha dovuto far ricorso alla *peer review* e ha vissuto “la VQR in termini competitivi (tra SSD o loro componenti, e tra individui)” optando perciò “per una valutazione generosa dei prodotti della loro Area e punitiva di quelli di aree considerate concorrenti”.

Sempre nelle *Conclusioni* si ragionerà di quali siano le strade per arrivare ad una maggiore sofisticazione, e su quali siano i rimedi più adatti a risolvere, o comunque affrontare, in futuro gli altri problemi emersi dalla VQR. Alla fine di questa introduzione ci limitiamo invece a riepilogare i principali elementi critici necessari all'interpretazione di questo *Rapporto*, che andrà letto tenendo presente:

1. Che ogni comparazione *diretta* tra aree e settori puramente bibliometrici e puramente *peer* è impropria;



2. Che le classifiche affidabili sono, ma solo in parte, quelle generali di Area (dove si può sperare che alcuni *bias* si compensino, una volta che esse siano presentate dividendo i macro-settori biblio da quelli *peer*), e *soprattutto* quelle per SSD o—se le si vuole creare—per gruppi di SSD affini per dimensioni e cultura della valutazione (o atteggiamento verso di essa). Solo all'interno di culture omogenee è stato infatti possibile tenere sotto controllo, per esempio attraverso la distribuzione dei prodotti a più *referee*, il *bias* soggettivo della valutazione *peer*;
3. Che quindi le comparazioni corrette tra SSD diversi, ma appartenenti alla stessa struttura (tanto Università che dipartimenti), sono quelle *indirette*, che mostrano cioè NON la % di Eccellenti o Limitati di questo SSD rispetto ad un altro SSD nella struttura data, ma le diverse posizioni che occupano gli SSD di una determinata struttura all'interno della classifica generale nazionale dei rispettivi SSD. Sono queste posizioni a poter e dover essere confrontate (per esempio: che posto occupa il SSD xyz dell'Ateneo x nella graduatoria nazionale del SSD in questione? E che posto occupa invece il SSD zyx dello stesso Ateneo? *Solo da questo confronto sarà possibile accertare il loro valore relativo*);
4. Che nella valutazione delle strutture complesse (con più aree o SSD) occorrerebbe pesare tali aree o tali SSD: in Area 11 un dipartimento con una prevalenza di piccoli SSD valutati in *peer* mostra per esempio strutturalmente un risultato migliore di un dipartimento dove predominano invece grandi SSD;

Che è interesse nazionale assicurare la promozione della parte migliore di *tutti* gli SSD, perché il paese è grande a sufficienza e ha un passato tale da poter ancora ritenere di coltivare l'intero spettro del sapere umano.

1.1 Descrizione dell'area

L'Area 11 è forse la più composita tra quelle CUN in cui sono tradizionalmente divise le discipline scientifiche in Italia. In essa convivono forti e diverse tradizioni disciplinari—Antropologia, Filosofia, Geografia, Pedagogia, Psicologia e Storia—che le conferiscono una grande ricchezza culturale e richiedono un approccio rispettoso delle loro specificità ed esigenze. Si tratta infatti di macro-settori a se stanti, diversi per metodologia, livello di internazionalizzazione, tipologia delle pubblicazioni, cultura della valutazione e atteggiamento verso di essa, come dimostra l'opzione a favore della bibliometria della Psicologia e di due SSD M-EDF, formalmente uniti alla Pedagogia. Nelle Tabelle si è quindi deciso di disaggregare il



subGEV più grande, presentando valori separati per Antropologia, Geografia, Scienze del Libro e del Documento (Archivistica-Biblioteconomia-Paleografia) e Storia, sia per rispettare le peculiarità proprie di queste discipline, sia perché si è ritenuto di fornire così al lettore dati più omogenei e interessanti.

I sei macro-settori sono a loro volta divisi in 38 SSD, e ora in 17 SC, che riuniscono al 10 giugno 2013 un totale di 4.749 soggetti valutati, dei quali 4.613 facenti capo alle università e i restanti agli enti di ricerca e ai consorzi. Fra essi si contano 9 dirigenti di ricerca, 23 primi ricercatori, 4 tecnologi, 1.914 ricercatori (di cui 1.826 facenti capo alle università), 1.392 professori associati, 1.406 professori ordinari (di cui 10 a tempo determinato) e 1 professore incaricato. Di essi 3.427 fanno riferimento alla sub-area non bibliometrica e 1.322 a quella bibliometrica (per i dettagli della loro ripartizione a livello di subGEV, SSD e numero di prodotti attesi, cfr. Tabelle 1.6-1.9).

All'interno della stessa Area 11, inoltre, vi sono SC, e persino SSD, come MSTO/08, composti riunendo vecchi SSD aboliti perché troppo piccoli, ma ancora caratterizzati da forti identità disciplinari. La complessità che caratterizza l'Area 11 nella sua interezza è così replicata anche al suo interno.

Non solo: il processo di valutazione ha dimostrato con chiarezza che vi sono SSD di altre aree che presentano notevoli affinità con quelli dell'Area 11, come è stato riconosciuto dalla richiesta che fosse l'Area 11 ad occuparsi della loro valutazione. Ciò è successo per gli SSD delle Storie di Africa, Asia e Americhe (SPS/05, 13 e 14), formalmente afferenti all'Area 14, e probabilmente il discorso si sarebbe potuto estendere alla Storia delle Relazioni Internazionali (SPS/06), che tanti legami ha con la storia moderna e contemporanea. Ma è successo anche con la Storia della Medicina (MED/02) e forse sarebbe dovuto succedere—almeno a livello della valutazione—anche con le storie delle altre scienze (fisica, matematica, chimica, psicologia ecc.), che hanno talvolta sofferto la valutazione bibliometrica e avrebbero forse potuto essere valutate associandole al SSD di Storia della Scienza e delle Tecniche (M-STO/05) presente nell'Area 11.

Il discorso potrebbe inoltre essere allargato alle storie del mondo classico, così vicine per tanti versi a quelle dell'Area 11, anche se in questo caso i motivi che spingono a un loro raggruppamento con le altre scienze dell'antichità sono forti. Questo non vuol dire negare la specificità di SSD che hanno fin qui, e con buone ragioni, trovato casa in altre aree CUN. Ma questi stessi SSD hanno spesso sofferto di accorpamenti di cui la valutazione ha messo in luce



l'irrazionalità, almeno ai suoi fini, suggerendo che le aree CUN potrebbero, sempre ai fini della valutazione, essere ridisegnate. Ciò potrebbe essere fatto accorpando SSD affini e oggi divisi tra esse (un esempio ovvio è quello della storia della scienza), ma anche all'occorrenza riconoscendo la piena autonomia di grandi macro-settori la cui riunificazione in un'unica Area non trova convincenti basi logiche e culturali.

1.2 La composizione del GEV

Il GEV 11 conta 38 componenti, compreso il suo presidente, il cui lavoro è stato validamente coadiuvato da quello di un assistente. Di regola il Direttivo ANVUR ne ha scelto i membri tra i professori ordinari che avevano fatto domanda come valutatori del Comitato di Indirizzo per la Valutazione della Ricerca (CIVR), poi selezionati in base alla qualità, quantità e internazionalizzazione della loro produzione scientifica (cfr http://www.anvur.org/index.php?option=com_content&view=article&id=102:messaggio-di-accompagnamento-liste-gev-15-dicembre-2011-it&catid=26&Itemid=188&lang=it). Le eccezioni sono state dovute alla necessità di introdurre opportuni correttivi per tener conto della tipologia e distribuzione geografica degli Atenei italiani, nonché per assicurare il necessario equilibrio di genere e una maggiore presenza internazionale.

L'ANVUR ha altresì cercato di fare in modo che tutti i SSD dell'Area fossero rappresentati nel GEV, tenendo tuttavia conto delle dimensioni molto variabili degli stessi SSD, tra cui ve ne sono alcuni che contano diverse centinaia, e altri poche decine, di aderenti. Idealmente, sarebbe stato inoltre opportuno evitare che singoli SSD fossero affidati a un solo membro GEV. Naturalmente, con 38 componenti e 33 SSD, trovare un equilibrio perfetto era impossibile ed è stato possibile assicurare la presenza di almeno due membri per SSD solo nei grandi SSD. Si è rimediato affiancando ai membri GEV che rappresentavano un solo SSD colleghi di SSD affini.

Si è così arrivati ad un GEV con 24 uomini e 14 donne, di cui 6 insegnano all'estero. Essi rappresentano 24 Atenei (ai quali vanno aggiunti 6 tra atenei ed enti di ricerca esteri), di cui 5 nel nordovest, 6 nel nordest, 7 nel centro, e 5 nel meridione e nelle isole. La Sapienza Università di Roma e l'Università cattolica "Sacro Cuore" sono gli Atenei col maggior numero di EV (rispettivamente quattro e tre). La composizione del GEV, e il numero di prodotti gestiti da ciascun componente, sono indicati nelle tabelle 1.10 e 1.11.



Il Presidente ha poi proceduto, come consentitogli, a strutturare il GEV per sub-GEV, ciascuno dotato di un coordinatore, cui è stata data ampia autonomia all'interno tuttavia di un dialogo costante. In particolare sono stati creati i sub-GEV di Antropologia, Geografia e Storia (coordinatore Francesco Benigno, 14 membri), di Filosofia (coordinatore Diego Marconi, 9 membri), di Pedagogia (coordinatore Giuseppe Bertagna, 6 membri) e di Psicologia (coordinatore Stefano Puglisi Allegra, 8 membri). Quando lo hanno ritenuto opportuno, i sub-GEV hanno tenuto—previa autorizzazione—riunioni separate, in cui sono stati affrontati i problemi specifici della discipline da essi raggruppate.

I sub-GEV hanno poi diviso tra i loro membri i prodotti da valutare, prodotti che hanno gestito in piena autonomia, nel rispetto delle specificità disciplinari ma attenendosi ai criteri decisi all'unanimità all'inizio del processo di valutazione, e discutendo e risolvendo gli eventuali problemi col presidente del GEV via via che essi si presentavano. Costante è stata inoltre la consultazione telematica, che ha visto lo scambio di migliaia di mail.

1.3 I tempi

Il GEV ha lavorato per 18 mesi, dal dicembre 2011 al maggio 2013. Nei mesi di gennaio e febbraio 2012 è stato discusso e poi approvato per via telematica il *Documento dei criteri per la valutazione dei prodotti di ricerca dell'Area 11*, pubblicato sul sito ANVUR il 29 febbraio 2012 (vedi Appendice B), elaborato in base alle discussioni tenutesi nella prima riunione plenaria del GEV, convocata a Roma il 16 dicembre 2011, di cui è stato a suo tempo reso pubblico il verbale. Contemporaneamente si è proceduto, in collaborazione con le Società e le Consulte dell'Area, al lavoro necessario alla redazione del Documento di lavoro *Per una classificazione della Riviste scientifiche di Area 11* (vedi Appendice B).

Dopo la pubblicazione dei criteri, il lavoro si è concentrato sulla selezione delle migliaia di *referee* necessari in un'Area che, in base alle previsioni, aveva bisogno di circa 24000 valutazioni *peer*.

A fine aprile 2012 tutto era pronto per cominciare il lavoro di valutazione, ma a questo punto una serie di slittamenti progressivi—dovuti a richieste successive di Atenei e CNR di avere più tempo per la selezione dei prodotti, e poi all'avvio delle Abilitazioni scientifiche nazionali—ha ritardato l'invio ai *referee* dei prodotti da valutare ai primi di settembre, provocando non pochi problemi, visto che era stato preannunciato ai *referee* stessi che avrebbero



avuto la possibilità di lavorare nei mesi estivi, quando gli impegni didattici e amministrativi sono minori, e su questa base si era ottenuto il loro consenso.

Una volta avviato il lavoro, questo è tuttavia proceduto con relativa speditezza, nonostante i non piccoli problemi rappresentati dalla gestione di migliaia di PDF e centinaia di *referee*, dalla necessità di dover far fronte ai rifiuti di questi ultimi e di provvedere alla sostituzione dei PDF difettosi (che richiedeva l'aiuto degli Atenei) ecc.

Nel corso di questi mesi alcuni sub-GEV si sono riuniti, altri hanno preferito gestire il lavoro attraverso il coordinamento telematico tra i gruppi di lavoro in cui si erano a loro volta suddivisi. In particolare il sub-GEV di Psicologia si è riunito per la prima volta il 16 dicembre 2011 in occasione della prima riunione del GEV presso il MIUR, per poi riunirsi ancora tre volte (27/03/2012, 04/07/2012, 18/01/2013), per discutere e definire aspetti cruciali nell'attuazione del processo di valutazione con particolare riferimento all'applicazione dei criteri specifici della Psicologia (vedi l'Allegato n. 2 dei criteri del GEV 11). Il 23/11/2012 si è tenuta invece una riunione dei coordinatori dei sub-GEV col presidente del GEV, alla presenza del coordinatore della VQR, riunione alla quale è intervenuto anche il presidente dell'ANVUR (cfr. tabella 1.10).

Il grosso del lavoro è stato terminato nel febbraio 2013, quando, il 1° del mese, si è tenuta la seconda e ultima riunione plenaria del GEV. Nelle settimane successive, in base alle decisioni prese nella riunione, sono stati affrontati e risolti i problemi posti dai prodotti con valutazioni discordanti, quelli sollevati dall'esclusione dalla valutazione di prodotti che non corrispondevano ai criteri stabiliti dal GEV e dalle relative penalizzazioni, e quelli legati alla necessità di valutare con la *peer review* un certo numero di prodotti bibliometrici per cui le matrici predisposte in precedenza e pubblicate nei criteri del GEV non avevano dato risultati conclusivi. Le valutazioni dei prodotti ed il rapporto finale sono stati infine approvati in una riunione plenaria tenutasi per via telematica il 20 giugno 2013.

2 La Valutazione dei “prodotti di ricerca”

2.1 I criteri di valutazione

Nel portare avanti la VQR l'Area 11 ha sempre tenuto fermo il principio del massimo rispetto della specificità, e quindi delle decisioni, dei macro-settori da cui essa è composta, ponendosi al tempo stesso il compito di assicurare che la valutazione venisse effettuata nel modo



più serio, rigoroso ed equilibrato possibile, cercando di neutralizzare ogni spinta o fattore potenzialmente capace di influire negativamente su quel rigore e quell'equilibrio.

Ciò è stato fatto applicando i criteri di valutazione elaborati dal GEV tra il dicembre 2011 e la fine del febbraio 2012, di cui si richiamano qui alcuni degli elementi principali.

In particolare il GEV ha seguito il metodo della *informed peer review* nella maggioranza dei suoi macro-settori, e in prevalenza quello bibliometrico nella Psicologia e nell'area biomedica dei SSD M/EDF. Questa *informed peer-review* è stata attuata con metodologia *single-blind* (data l'impossibilità di usare quella *double-blind* su prodotti già pubblicati), ed è stata applicata sia agli articoli su rivista, che alle monografie e ai capitoli di libro, con la differenza che nel caso degli editori non è stato prodotto alcun *rating*, e si è chiesto ai *referee* di tener anche conto—quando possibile—delle eventuali recensioni su riviste, specie se di fascia A o internazionali, e della qualità della sede di pubblicazione, indicata dall'esistenza di un comitato editoriale e di procedure trasparenti di lettura per decidere sulla pubblicazione, dalla diffusione a livello nazionale e internazionale dei prodotti dell'editore e da ogni altro elemento atto a fornire indicazioni utili sulla qualità e impatto dell'opera.

Per *peer-review* si è intesa la valutazione di pari non dal punto di vista del rango accademico formale, ma della competenza scientifica. Ogni prodotto è stato affidato a due revisori esterni (cfr. *infra* per la loro selezione), a eccezione dei saggi pubblicati su riviste internazionali di fascia A segnalate dai vari SSD e delle monografie pubblicate dalle maggiori case editrici internazionali, per cui gli EV (Esperti valutatori, vale a dire i membri GEV) potevano proporre che fosse sufficiente una sola valutazione.

E' stata posta la massima cura nello scegliere revisori che fossero studiosi e specialisti autorevoli, scientificamente qualificati e attivi nel periodo della VQR, delle discipline cui appartenevano i prodotti da valutare. Si è cercato altresì di fare in modo che fossero dotati di imparzialità, rigore, equilibrio e senso dei propri limiti. Il possesso di un dottorato di ricerca, o il ruolo docente (emeriti inclusi), sono stati considerati requisiti minimi.

L'individuazione dei revisori è stata di regola fatta separatamente da gruppi di lavoro composti da almeno due EV diversi, e più spesso da tre o anche quattro EV, e si è avuto cura di inviare i prodotti di ciascun soggetto valutato ad almeno quattro revisori diversi, e di preferenza a sei, facendo attenzione ai casi di incompatibilità e ai possibili conflitti di interesse (tra cui per esempio parentela, appartenenza allo stesso Ateneo, corrente partecipazione agli stessi progetti di ricerca ecc.). Per minimizzare questi problemi, e per dare visibilità internazionale ai prodotti



valutati, laddove possibile sono stati privilegiati revisori operanti al di fuori dei confini nazionali e attivi nelle migliori università e centri di ricerca internazionali.

I revisori sono stati informati dei criteri elaborati dal GEV, e hanno ricevuto insieme ai prodotti da valutare indicazioni di vario tipo, tra cui quelle relative ai *rating* delle riviste, cui però non sono stati tenuti a uniformarsi. Nella valutazione essi hanno seguito le regole e le procedure indicate nella “scheda valutazione prodotti” predisposta sulla base delle indicazioni ANVUR e poi in parte modificata dal GEV.

I membri del GEV sono stati usati come revisori, ma—salvo casi eccezionali, autorizzati dal presidente—nessun prodotto è stato affidato a due membri del GEV. I revisori dei prodotti dei componenti del GEV sono stati scelti dai coordinatori dei sub-GEV corrispondenti; il presidente del GEV ha scelto i revisori esterni dei lavori scientifici dei coordinatori di sub-GEV; il coordinatore della VQR ha indicato i revisori esterni dei lavori del presidente del GEV.

Per i casi di valutazioni non convergenti dei revisori *peer*, i sub-GEV hanno creato al loro interno gruppi di consenso con il compito di proporre al GEV il punteggio finale. La responsabilità della valutazione conclusiva è infatti dell'intero GEV, non del sub-GEV e mai del componente che appartiene a uno specifico settore scientifico disciplinare. Per i prodotti con valutazioni radicalmente divergenti (E-L) è stato acquisito il parere di un terzo revisore esterno, salvo casi eccezionali.

2.2 La classificazione delle riviste

Il GEV 11 ha lavorato alla classificazione delle riviste “correnti” (definite in via provvisoria, salvo prevedere per il futuro criteri più stringenti, come tutte le riviste che avevano pubblicato a fine 2011 almeno un numero dell'annata 2010) tra il dicembre 2011 e il febbraio 2012. Il *rating* finale è stato articolato in due fasce A e B, ritenendo implicitamente la C come composta residualmente dalle altre riviste scientifiche proprie di ciascun SSD. Il lavoro, descritto in dettaglio nel già citato documento *Per una classificazione delle Riviste scientifiche di Area 11*, è stato reso possibile dalla collaborazione delle Consulte e delle Società di ogni SSD. E' stata una collaborazione non priva di momenti di confronto, ma sempre aperta, trasparente e improntata al comune desiderio di tenere fermamente la valutazione nelle mani della comunità scientifica.

In particolare, seguendo le indicazioni ricevute dall'ANVUR, la procedura che ha portato



a questo risultato è stata svolta in quattro tappe:

1. Si sono chiesti alle Società e alle Consulte degli elenchi divisi in due fasce (A e B) delle riviste italiane, e nel caso internazionali, rilevanti per ciascun SSD, nonché delle riviste intersettoriali e interdisciplinari per esso più importanti. Per ciascuna fascia sono stati indicati dei tetti quantitativi. Le riviste nate dopo il 2008 hanno ricevuto una valutazione provvisoria, come segnalato in ciascun caso;
2. Questi elenchi sono stati sottoposti a dei revisori italiani e stranieri (in genere tre per elenco), scelti tra specialisti delle discipline che avevano trasmesso gli elenchi e selezionati tra gli studiosi più autorevoli e degni di stima;
3. Il risultato dei referaggi è stato poi sottoposto alle Società e alle Consulte, che hanno fatto le loro controdeduzioni;
4. Gli elenchi così rivisti, unificati per grandi macro-settori in modo che le riviste selezionate come A o B fossero tali per tutti, sono stati poi presentati ai sub-GEV, e infine approvati dal GEV 11 nel suo complesso.

Tale procedura, approvata dall'ANVUR, ha dato risultati soddisfacenti, come dimostrano le poche domande di revisione, in genere avanzate da riviste escluse dalla fascia A, nella finestra aperta a tal fine nel maggio 2012. Soltanto 18 riviste, sulle centinaia e centinaia valutate, hanno allora contestato la scelta fatta, che è stata riconsiderata, modificando il giudizio precedente in sette casi.

La divisione adottata nel *rating* delle riviste in italiane e internazionali (tra le quali naturalmente si possono trovare anche riviste pubblicate in Italia) registra una differenza oggettiva. Le riviste internazionali, che si rivolgono agli specialisti di tutto il mondo in questa o quella disciplina, non possono essere confrontate con quelle puramente nazionali, prodotte da comunità più piccole e meno variegata, ancorché spesso di qualità eccellente. Tuttavia, data la peculiarità della produzione in campo umanistico, ancora spesso legata a una dimensione prevalentemente nazionale, sembrava errato e ingiusto relegare in B tutte o quasi le riviste italiane, che per ovvi motivi non raggiungono le dimensioni e non hanno la vitalità di riviste che sono alimentate dalla ben più vasta *république des lettres* internazionale.

Vista la libertà lasciata ai revisori, è difficile giudicare dell'utilità e dell'uso dei *rating* nel processo effettivo di valutazione. L'impressione è che vi sia stata una notevole variabilità tra SSD diversi, con un rapporto più stretto tra qualità della rivista e classificazione in E-L degli articoli in essa pubblicati negli SSD più internazionalizzati. In ogni caso si rimanda alla Sezione



2.7, infra, per un'esposizione sistematica di tale rapporto nella generalità degli SSD.

I risultati della VQR hanno tuttavia pienamente confermato la differenza sostanziale nella qualità dei prodotti pubblicati tra grandi riviste internazionali (alcune delle quali basate in Italia) e le riviste nazionali, componendo una graduatoria che vede le prime saldamente al primo posto, le seconde a un gradino più in basso, e le riviste B di fatto equiparabili all'insieme delle riviste scientifiche (vedi infra 2.6 per un'analisi più dettagliata).

Un problema particolare, che rimanda a quello della difformità delle aree CUN rispetto alla valutazione, si è posto per i SSD, come per esempio Storia della Medicina, formalmente inseriti in aree bibliometriche, e quindi privi di *rating* di riviste, ma che hanno poi sottoposto buona parte della loro produzione scientifica alla *peer review*. Si tratta di un problema che era impossibile, per ragioni formali, risolvere nel corso della VQR, ma che andrà affrontato e risolto in futuro.

2.3 Il processo di valutazione

Il processo di valutazione nell'Area non ha sollevato conflitti significativi. Anche se naturalmente e per fortuna non sono mancate le discussioni, specie su temi specifici a singoli sub-GEV, tutte le decisioni, a cominciare da quelle relative alle questioni principali, sono sempre state prese all'unanimità.

Come previsto dai criteri dell'Area nessun SSD è stato affidato ad un solo EV. Per i singoli SSD sono stati costituiti gruppi di lavoro di grandezza variabile (3-4 persone) a secondo del sub-GEV e della grandezza del SSD in questione. Il presidente si è in genere limitato a monitorare lo svolgimento di un processo che è stato nel suo insieme serio e corretto, intervenendo solo quando necessario o su richiesta dei coordinatori di sub-GEV o dei capifila dei gruppi di lavoro.

I lavori valutati di Area 11 sono stati complessivamente 13.152, un dato che comprende 450 prodotti sottoposti ad altre Aree e mandati da queste ultime in valutazione al GEV 11, a fronte di 13.487 prodotti teoricamente attesi.

I prodotti mancanti (vale a dire attesi e non spediti né all'Area 11 né ad altre Aree, a segnalare situazioni di parziale o totale inattività) sono stati 335, pari al 2,5% del totale, di cui 265 nell'area non bibliometrica e 70 in quella bibliometrica (cfr. Tabelle 2.1-2.2).



10.905 lavori, pari al 82,9% del totale, sono stati valutati in *peer*. 2.146 lavori, pari al 16,3% del totale, appartenenti al macro-settore della Psicologia o agli SSD M-EDF, sono stati invece valutati bibliometricamente.

Al GEV-11 sono stati inoltre conferiti 450 prodotti proposti da ricercatori afferenti ad altre aree, mentre 456 prodotti sono stati proposti ad altri GEV da ricercatori afferenti all'area 11. Di essi 450 sono stati effettivamente valutati in altre aree: 26 in area 1, 2 in area 2, 1 in area 3, 97 in area 5, 153 in area 6, 3 in area 7, 15 in area 8, 20 in area 9, 41 in area 10, 8 in area 12, 26 in area 13 e 58 in area 14.

Inoltre, come si è detto, il GEV-11 ha gestito la valutazione *peer* di alcuni interi SSD di altre aree, e precisamente di quello di Storia della Medicina (MED/02) in Area 6 e di quelli di Storia e Istituzioni delle Americhe (SPS/05), Storia e Istituzioni dell'Africa (SPS/13) e Storia e Istituzioni dell'Asia (SPS/14) in Area 14. Ciò ha comportato la gestione di un totale di ulteriori 272 prodotti – di cui 89 di Area 6 e 183 di Area 14, dei quali 61 di Storia e Istituzioni delle Americhe (SPS/05), 72 di Storia e Istituzioni dell'Africa (SPS/13) e 50 di Storia e Istituzioni dell'Asia (SPS/14) – che sono poi “ritornati” alle rispettive Aree e dunque non compaiono nelle statistiche relative all'Area 11.

A. Macro-settori con *peer review*

Nella *peer review*, ogni prodotto è stato assegnato a uno dei sub-GEV in base al suo SSD e poi ripartito tra gli appositi gruppi di lavoro che sulla base delle competenze dei *referee* disponibili hanno cercato di assegnare i tre prodotti di ciascun autore a sei *referee* diversi, o dove questo era impossibile (per esempio nei campi di studio molto specialistici) ad almeno quattro. I prodotti di ricerca dei membri GEV e del Presidente GEV sono stati valutati seguendo le regole previste dalla VQR (cfr 2.2 supra).

In media, ciascun EV è stato responsabile di 307 lavori inviati in *peer review*; a livello di sub-GEV le medie sono rispettivamente 338 (Antropologia, Storia, Geografia); 324 (Filosofia); 345 (Pedagogia), e di 195 nel caso della Psicologia, che ha però valutato bibliometricamente buona parte dei suoi prodotti (calcolando anche i prodotti valutati in bibliometria, ogni componente del subGEV di Psicologia è stato infatti responsabile di 390 prodotti).

I gruppi di lavoro hanno prima esaminato i prodotti per individuare ed escludere i “prodotti non valutabili” (lavori appartenenti a tipologie escluse dalla VQR, con allegati mancanti o non adeguati per la valutazione, pubblicati in anni precedenti o successivi al periodo



di riferimento della VQR, lavori non pubblicati, lavori senza alcun contenuto di ricerca come libri di esercizi per il triennio, ecc.). In tutti gli altri casi i prodotti sono stati inviati alla valutazione esterna.

I criteri seguiti nella selezione dei *referee* da utilizzare sono già stati ricordati in 2.1. Ispirandosi ad essi, ogni EV ha avuto la possibilità di proporre dei nomi; le proposte dei singoli EV sono state fatte circolare a livello di sub-GEV; e ogni EV ha avuto la possibilità di esporre direttamente al presidente (e quindi non pubblicamente) eventuali dubbi motivati relativi ad alcune proposte. Il presidente ha svolto in questi casi indagini supplementari e ha alla fine approvato, per sub-GEV e in accordo coi relativi coordinatori, la lista finale dei *referee*, garantendone quindi anonimato (nel suo complesso tale lista non ha mai circolato, e nella sua interezza essa è nota al solo presidente e all'assistente del GEV).

La lista così composta è stata integrata, quando ciò si è reso necessario per carenza di *referee* specialisti di singoli campi di ricerca, con nomi tratti dalla banca dati revisori del MIUR e dall'elenco delle autocandidature ricevute dal GEV. Questi, pochi, casi di integrazione sono stati sottoposti al parere dei coordinatori di sub-GEV o degli EV competenti per SSD. Altre poche integrazioni sono state effettuate nel corso del processo di valutazione, quando ci si è trovati di fronte a prodotti che richiedevano competenze non presenti nell'elenco dei *referee* selezionati nei primi mesi del 2012, o non utilizzabili a causa per esempio di conflitti di interesse.

Al momento dell'assegnazione dei prodotti ai *referee*, l'indipendenza dei revisori è stata garantita prestando attenzione alle sedi di lavoro, alle collaborazioni di ricerca ecc. Nei casi in cui il revisore abbia evidenziato la presenza di conflitti di interessi non noti preventivamente al GEV, una sua non competenza sullo specifico lavoro, o la mancata disponibilità per motivi di tempo, il gruppo di lavoro o l'EV che aveva proceduto all'assegnazione ha provveduto a modificarla e assegnare nuovamente il lavoro a un altro revisore.

I revisori hanno fornito una valutazione qualitativa dei prodotti di ricerca compilando *on line* una scheda di valutazione (riportata in Appendice 4) che ha sollevato problemi cui si è già accennato e che verranno ripresi nelle *Conclusioni*. Essa era costituita da tre domande tendenti a valutare le caratteristiche di rilevanza, originalità e internazionalizzazione previste dal bando e da un campo libero per un breve eventuale giudizio finale. Il punteggio di sintesi (ad ogni domanda si poteva assegnare un punteggio da 1 a 9) è stato ottenuto automaticamente dal CINECA sommando i punteggi delle tre risposte, e attribuendo una classe di merito complessiva.



Dato l'elevato numero di revisori stranieri, la scheda di valutazione è stata predisposta sia in italiano che in inglese.

La risposta dei colleghi italiani e stranieri cui è stato chiesto di agire da *referee* è stata nel complesso buona. Il GEV 11 ha potuto avvalersi di un totale di 1.442 revisori esterni, di cui 1.076 italiani e 366 stranieri, dei quali oltre duecento (fra cui 25 stranieri) hanno accettato di contribuire alla valutazione *peer* di oltre 25 prodotti (cfr. Tabella 2.3). Come si è detto, pochissimi hanno rifiutato l'incarico per contrarietà alla valutazione stessa e coloro che l'hanno accettato hanno operato in maniera collaborativa e impegnata. I casi di "collaborazione ostativa" (per esempio l'assegnazione di tutti i prodotti alla sola classe E) sono stati rari e i sub-GEV hanno raccolto pareri nel complesso positivi sulla VQR da parte di *referee* stranieri oltre che da molti italiani coinvolti, nonché commenti interessanti sulla qualità dei prodotti, giudicata spesso eccezionalmente eterogenea, dall'eccellente al pessimo.

Tuttavia, come ha notato un EV, almeno in alcuni SSD se "la disponibilità è stata in genere molto buona da parte dei colleghi italiani interni al sistema universitario, in termini di puntualità e anche di accettazione di prodotti 'sgradevoli' e numerosi, meno buona è risultata la disponibilità degli esterni e degli stranieri". Come hanno sottolineato tutti i coordinatori di sub-GEV, molti *referee* stranieri hanno per esempio accettato in prima battuta, ma hanno poi ritirato la loro disponibilità, probabilmente a causa dei già ricordati ritardi di avvio, che hanno impedito di utilizzare i più tranquilli mesi estivi per la valutazione dei prodotti e per le difficoltà connesse alla valutazione di una produzione che è risultata essere, forse più di quanto non fosse nelle attese, in lingua italiana. Vale la pena di aggiungere che in alcuni SSD, contrariamente alle attese, i *referee* stranieri si sono mostrati più generosi degli italiani. Essi inoltre hanno rifiutato o lasciato inevasa una maggiore percentuale di richieste di revisione rispetto a quanto hanno fatto i revisori interni al sistema universitario italiano: se, in media, questi ultimi hanno rifiutato o lasciato inevaso meno del 20% delle richieste loro pervenute, i *referees* stranieri hanno in media rifiutato o lasciato inevaso quasi il 40% delle richieste loro pervenute (cfr. Tabelle 2.4-2.6, per una panoramica della situazione anche a livello dei singoli SSD). Vi sono stati poi problemi con prodotti "difficili", rifiutati anche da più di sei valutatori, che hanno costretto i gruppi di lavoro e l'assistente a trovare soluzioni talvolta non facili.

Tuttavia, i problemi maggiori sono derivati dalla già ricordata forte componente soggettiva della valutazione *peer*. Come ha fatto rilevare un coordinatore di subGEV, "un esame delle valutazioni, e in particolare delle motivazioni sintetiche finali, mostra che *referee* diversi hanno spesso interpretato le categorie [da Eccellente a Limitato]... in modo diverso. Alcuni casi



tipici: un buon manuale, o un buon prodotto semi-divulgativo, è stato valutato E o B (in quanto buono) ma anche L (in quanto manuale). L'originalità è stata ritenuta dirimente da alcuni [revisori], ma non da altri. E l'internazionalizzazione è stata a volte interpretata come 'accessibilità alla comunità internazionale' (per cui un prodotto in lingua italiana è stato per ciò stesso valutato poco relativamente al terzo parametro...), altre volte come 'al livello della produzione media internazionale', altre volte ancora come 'apertura alla letteratura internazionale' (la pubblicazione non cita solo autori italiani), con risultati di valutazione molto diversi”.

Una scheda meglio articolata, di cui si discuterà nelle *Conclusioni*, avrebbe probabilmente ridotto questi problemi, cui si sono aggiunti quelli derivati da una misura che pure sembrava, ed era, auspicabile. L'aver fatto assegnare i prodotti di un singolo autore a EV diversi, onde correttamente evitare di affidare il suo “destino” a un solo giudice, ha infatti impedito agli EV di considerare la produzione di questo autore nel suo complesso, e quindi di rilevare i limiti relativi ad una possibile produzione monotematica, tipica dei casi in cui insieme ad una monografia un autore ha sottoposto alla VQR degli articoli che ne anticipavano, o sintetizzavano, i risultati (non si pensa naturalmente qui a chi ha dedicato lunghi anni allo studio di aspetti importanti e diversi di problemi o autori complessi). Il sistema garantista adottato non ha quindi consentito di cogliere queste duplicazioni e sarebbe perciò opportuno, pur salvaguardando quelle garanzie, trovare per il futuro—come suggerito da un EV—il modo di valutare e premiare la varietà degli studi, “ovvero di ponderare diversamente chi ha presentato tre prodotti largamente coincidenti e chi ha presentato tre studi di oggetto e campo diverso”.

Altri problemi sono infine stati causati dalla scarsa presenza in Italia di una cultura della valutazione, anche tra *referee* pure selezionati con cura. Come dimostra la Tabella 2.7 alcuni revisori hanno assegnato, se non sempre la stessa classe, ventagli molto limitati di classi: un comportamento dal genere è stato tenuto da una media del 21% degli 830 revisori, nominati dal GEV 11, che abbiano effettuato almeno 8 revisioni per SSD che hanno fatto interamente ricorso alla *peer review*. (Siccome pochi revisori hanno effettuato almeno quel numero minimo di revisioni per SSD che facevano principalmente ricorso alla bibliometria, un dato analogo calcolato su di essi avrebbe avuto scarso significato, e dunque non viene prodotta una tabella analoga per la sub-area bibliometrica). Si sono registrate in ogni caso differenze molto significative tra i vari SSD: in sette infatti il dato è stato superiore alla media, e in tre l'ha superata di gran lunga.



Da questo punto di vista la VQR è stata anche un'importante esperienza di referaggio in una comunità che almeno in parte delle discipline umanistiche non vi era abituata. Non va dunque sottovalutata la sua ricaduta in termini di diffusione di una “cultura della valutazione”, basata su standard minimi condivisi. Essa lascia in eredità al futuro una banca dati di *referee* filtrati in base al loro operato ed equilibrio, costruita rivedendo le liste composte a inizio 2012 in base all'esperienza fatta nei mesi successivi.

Nelle condivisibili parole di un EV, “è tuttavia possibile affermare che il referaggio sia stato in sostanza equilibrato, e che abbia rispettato, per così dire, ‘i valori in campo’, almeno come trend generale, malgrado cadute di tono che non hanno contaminato la validità del processo nel suo insieme”, anche per il pronto intervento degli EV, dei gruppi di lavoro, dei sub-GEV e, quando necessario, del presidente, nel correggere le distorsioni via via che queste si manifestavano (per esempio monitorando l'operato di *referee* che assegnavano i prodotti quasi esclusivamente in E o in L, oppure utilizzavano quasi esclusivamente una classe intermedia di valutazione). A riprova di ciò è per esempio possibile indicare il fatto che la VQR “ha premiato sia accademici *senior*, sia quelli *junior*, mostrando in questo modo un altro aspetto della sua utilità e, se vogliamo, democraticità” (vedi a proposito i dati nelle Tabelle 2.8-2.14, che riportano la distribuzione delle valutazioni per ciascuna fascia di docenza, divisi per SSD. Nel caso di lavori con più di un autore, ovviamente non è facile distinguere il contributo di un singolo autore, particolarmente nel caso in cui gli autori siano molti; nella tabella 2.14, riferita alla sub-area bibliometrica, dove questo tipo di prodotti sono più frequenti, si è supposto che ciascuno degli autori abbia contribuito nella stessa misura alla stesura del lavoro. In ogni caso, simili statistiche vanno interpretate con cautela, e servono per evidenziare le tendenze che emergono tra gruppi consistenti di ricercatori (strutture, dipartimenti, SSD, ecc.), non per analisi o valutazioni di casi singoli).

Ricevute le schede di valutazione dai valutatori, il GEV ha proceduto alle valutazioni di sintesi costituendo tanti Gruppi di Consenso quanti erano gli SSD di Area 11 (due altri sono stati costituiti *ad hoc* per i quattro SSD esterni all'Area di cui il GEV-11 ha gestito la valutazione *peer*). Questi gruppi hanno definito la valutazione di sintesi all'unanimità oppure ricorrendo—in caso di valutazioni fortemente discordanti, e in particolare E-L—a un terzo revisore. In caso di conflitto sulla valutazione finale del prodotto, i gruppi sono stati integrati dai coordinatori di sub-GEV o dal presidente del GEV.

I casi di conflitto radicale di valutazione (E-L) affrontati sono stati 417, con percentuali costantemente al di sotto del 2% nei SSD bibliometrici (tranne che in un caso in cui hanno



toccato il 5,5% del totale, e che forse non casualmente è quello del settore che meno ha fatto ricorso, tra quelli del subGEV di Psicologia, alla valutazione bibliometrica) e variazioni abbastanza significative tra gli SSD che si sono avvalsi unicamente della *peer review* (si va dal 1 al 7,5% circa del totale dei prodotti), a conferma delle diverse culture valutative e della variabilità dello strumento *peer* (cfr. Tabelle 2.15-2.16).

Nonostante le significative variazioni tra SSD, la relativamente piccola percentuale di prodotti con valutazioni radicalmente diverse (E-L) conferma ulteriormente l'affidabilità della VQR, specie *a livello di SSD*, dimostrata dal fatto che i prodotti con valutazioni concordanti o sostanzialmente tali (cioè con una sola classe di differenza tra quelle espresse dai due *referees*: E-B, B-A, A-L) sono quasi i quattro quinti di quelli valutati nei SSD che non hanno fatto ricorso alla bibliometria. In media infatti, poco più del 37% dei prodotti in questione hanno ricevuto valutazioni concordanti, e poco più del 40% valutazioni con una sola classe di differenza tra quelle espresse dai due *referees* interpellati. Anche in questo caso si sono registrate significative diversità tra i singoli SSD, ma nei settori non bibliometrici la percentuale di prodotti con valutazioni concordanti o sostanzialmente tali non è mai scesa sotto il 70%, superando la media in dodici casi (cfr. Tabelle 2.17-2.18).

Un problema particolare è stato presentato prima dalla definizione e poi dalla individuazione dei lavori “non valutabili”. Il bando VQR, che prevedeva linee guida di carattere generale circa la tipologia e l'ammissibilità dei prodotti, è stato interpretato in modo diverso dai singoli atenei e ricercatori. Inoltre, in alcuni casi (marginali) capire quale fosse la tipologia di pubblicazione non era semplice. Infine, non conoscendo *ex-ante* la tipologia dei lavori presentati alla VQR, i criteri stessi di valutazione del GEV 11 non avevano previsto tutte le categorie che ci si è poi trovati ad affrontare una volta che si è trattato di esaminare migliaia di prodotti.

Si è quindi proceduto con grande cautela. I lavori “non valutabili” sono stati evidenziati in prima battuta dagli EV, e ad essi si sono poi aggiunti quelli giudicati tali dai *referee*. I singoli casi sono stati poi esaminati dai gruppi di consenso, e la lista finale è stata rivista dai coordinatori sub-GEV e infine, e più volte, dal presidente. Nel caso di impossibilità di valutazione derivante da assenza o malfunzionamento dei PDF, è stato a più riprese chiesto alle strutture di mandarne di corretti, come è infatti poi avvenuto nella maggioranza dei casi. Nel caso di impossibilità di valutazione derivante dalla tipologia del prodotto, o da sue ambiguità, le motivazioni della problematicità del prodotto sono state inviate alle strutture per consentire le opportune verifiche e controdeduzioni, che sono state poi esaminate dai coordinatori di sub-GEV e dal presidente



ricorrendo quando opportuno alla collaborazione di EV competenti. Alla fine, il GEV ha considerato non valutabili 101 lavori.

B. *Psicologia e M-EDF*

Per quanto riguarda la Psicologia, il sub-GEV ha deciso e subito reso noto (come specificato nel febbraio 2012 nell'Allegato 2 ai criteri del GEV 11) che la scelta di ricorrere alla valutazione bibliometrica comportasse l'attribuzione di un valore diverso ai prodotti a diffusione internazionale rispetto a quelli a diffusione nazionale. Ciò è stato regolamentato nei criteri della Psicologia, prevedendo al contempo dei meccanismi di tutela delle aree meno consolidate sul piano della produzione internazionale.

Nel caso dei prodotti di Psicologia non sottoponibili alla bibliometria e quindi valutati in *peer review*, seguendo la procedura sopra descritta, i criteri adottati hanno consentito a quelli a diffusione internazionale (privi di IF) di accedere alle fasce superiori, e a quelli a diffusione nazionale, di essere collocati fino e non oltre la fascia degli Accettabili. Anche in questo caso, i prodotti *peer*, suddivisi per SSD, sono stati presi in carico da terne comprendenti due componenti del sub-GEV con competenza specifica nel SSD e il coordinatore del sub-GEV. La valutazione è stata poi gestita dai responsabili di ogni SSD che hanno assegnato i prodotti ai *referee* disponibili sulla base della loro competenza specifica nell'area tematica di riferimento del prodotto. Il processo di valutazione per la parte bibliometrica si è svolto attraverso la valutazione dei prodotti per *Subject Categories* e degli indici di citazione (ISI, Scopus) con l'ausilio della *peer* per i prodotti IR (vedi l'algoritmo bibliometrico nei già ricordati criteri del GEV 11).

Sono stati valutati in totale 3603 prodotti, di cui circa il 59% in bibliometria (cfr. Tabella 2.19). Ai prodotti M-EDF valutati in bibliometria (pari al 78% del totale di 327) sono state applicate le regole seguiti dalle aree bibliometriche. I prodotti assegnati alla *peer review* hanno invece seguito le procedure descritte in A.

2.4 I risultati della valutazione

Nella lettura e soprattutto nell'utilizzo di tutte le Tabelle riferite ai risultati della valutazione occorre sempre ricordare che i confronti diretti tra Aree e SSD sono impropri. Si rimanda a proposito a quanto detto alla fine dell'introduzione, e precisamente ai punti 1-3: i confronti corretti sono quelli *indiretti*, che confrontano le posizioni occupate da questo o quel SSD di un determinato Ateneo o dipartimento nella classifica generale del SSD stesso.



In aggiunta a quelli già menzionati nei paragrafi precedenti, in allegato vengono quindi forniti i seguenti dati:

- a. la somma dei punteggi ottenuti, la distribuzione dei prodotti nelle 4 classi di merito (E-B-A-L) e il voto medio per subGEV (o sottoinsiemi omogenei per quello di Storia, Geografia e Demotnoantropologia) e per SSD (cfr. 2.20-2.23). Questi dati confermano la variabilità, già ricordata nell'introduzione, emersa nei risultati della valutazione *peer* tra SSD appartenenti allo stesso macro-settore e persino allo stesso Settore concorsuale. Le variazioni esistenti anche tra SSD bibliometrici appartenenti al medesimo Settore concorsuale inducono a pensare che il dato possa essere influenzato anche da una disomogeneità di fondo di questi nuovi Settori.
- b. la somma dei punteggi ottenuti, la distribuzione dei prodotti nelle 4 classi di merito (E-B-A-L) e il voto medio per ciascuna tipologia di prodotto (Tabelle 2.24-2.25);
- c. il confronto tra Punteggi ottenuti e distribuzione dei prodotti conferiti non penalizzati nelle classi di merito (Eccellente -E-, Buono -B-, Accettabile -A, Limitato -L-) per modalità di valutazione (*peer review* o bibliometria), per la sub-area bibliometrica (Tabella 2.26).

2.5 I prodotti conferiti e la classificazione delle riviste

L'analisi dei dati relativi al confronto tra le classificazioni di riviste fatte *ex ante* dal GEV-11 e le valutazioni emesse *ex post* dal GEV stesso sui singoli prodotti rientranti nella categoria "articolo su rivista" conferma e precisa quanto ipotizzato nel già ricordato Documento di lavoro stilato all'inizio della VQR:

1. Nel loro insieme le riviste A ottengono risultati migliori delle rimanenti: calcolando fra di esse anche quelle internazionali (siano esse A o B – ciò è ininfluenza, alla luce di quanto detto nel Documento di Lavoro circa la differenza oggettiva tra riviste nazionali ed internazionali), emerge che esse hanno in media il 24% di Eccellenti contro una media del 17%, calcolata sul totale di tutti gli articoli valutati da SSD non bibliometrici, e il 73% di prodotti Eccellenti o Buoni contro una media del 61%, calcolata sul totale di tutti gli articoli valutati da SSD non bibliometrici (cfr. Tabella 2.27). Tale differenza probabilmente si accentuerebbe se si scorporassero le riviste dei piccoli SSD, che hanno altissime percentuali di prodotti eccellenti e tendono quindi ad appiattire il dato generale;



2. Tale differenza diventa davvero notevole se, ricordando quanto si è detto circa la differenza tra A internazionali e nazionali, si considerano solo le riviste internazionali; i prodotti pubblicati su queste ultime hanno in media il 40% di Eccellenti, con punte del 44% per le riviste A internazionali, e poco più del 10% di Accettabili o Limitati a fronte di una media del 18%; anche le B internazionali, peraltro, registrano una simile percentuale di prodotti Accettabili o Limitati e un 33% di Eccellenti (quasi il doppio della media del 17% calcolata sul totale di tutti gli articoli valutati da SSD non bibliometrici). Inoltre, su queste stesse riviste si concentrano il 32% di *tutti* gli articoli su rivista valutati come Eccellenti e solo il 2% di *tutti* gli articoli su rivista valutati come Limitati, benché i prodotti pubblicati su di esse costituiscano poco più del 14% del *totale degli articoli su rivista* valutati da SSD non bibliometrici (cfr. Tabelle 2.28-2.30).
3. Le riviste B NON presentano dati significativamente differenti da quelli che si registrano per le C (che sono in effetti le riviste scientifiche non classificate) e sembra quindi ragionevole aggregare le une e le altre in una categoria “residuale”; essa (B+C) registra poco più del 10% di prodotti Eccellenti e oltre il 27% di Limitati, a fronte di medie, calcolate sul totale di tutti gli articoli valutati da SSD non bibliometrici dell’area 11, pari rispettivamente al 17% e al 18%; (cfr. Tabella 2.31, Appendice A1);
4. I dati prefigurano quindi una classificazione di questo tipo: A (internazionali, incluse quelle pubblicate in Italia purché si rivolgano agli specialisti di tutto il mondo in questa o quella disciplina); B (eccellenti nazionali, ex A); C (residuale, inclusiva delle ex B nazionali);
5. La classificazione di area è piuttosto robusta. Non solo le riviste A e internazionali hanno il 24% di Eccellenti contro una media del 17%, ma il 70% degli articoli ritenuti Eccellenti, vale a dire 275 su un totale di 393, sono stati pubblicati su riviste classificate A o internazionali (cfr. Tabella 2.32). In particolare, alla luce dei dati, i *rating* di subGEV più robusti sono quelli di Storia e di Filosofia, seguiti in ordine di robustezza da quelli di Geografia, Archivistica-Biblioteconomia-Paleografia, Antropologia e Pedagogia.

In particolare, nel caso della Storia, l’84% degli articoli valutati Eccellenti sono stati pubblicati su riviste classificate A (comprese le internazionali), che hanno registrato una percentuale di valutazioni Eccellenti del 24% contro una media (calcolata sugli articoli su rivista valutati dagli SSD di Storia) del 14% e una di valutazioni Limitato del 7% contro una media di oltre il 20%. Nel caso della Filosofia, il 73% degli articoli valutati Eccellenti sono stati pubblicati su riviste classificate A (comprese le internazionali), che hanno registrato una percentuale di



valutazioni Eccellenti del 31% contro una media (calcolata sugli articoli su rivista valutati dagli SSD di Filosofia) del 23%, e una di valutazioni Limitate del 7% contro una media del 15%. Anche nel caso della Geografia, il 60% degli articoli valutati Eccellenti sono stati pubblicati su riviste classificate A (comprese le internazionali), che hanno registrato una percentuale di valutazioni Eccellenti del 16% contro una media (calcolata sugli articoli su rivista valutati dagli SSD di Geografia) del 11%, e una di valutazioni Limitate del 14% contro una media del 22%. Nel caso delle Scienze del Libro e del Documento (Archivistica-Biblioteconomia-Paleografia), il 59% degli articoli valutati Eccellenti sono stati pubblicati su riviste classificate A (comprese le internazionali), che hanno registrato una percentuale di valutazioni Eccellenti del 23% contro una media (calcolata sugli articoli su rivista valutati dagli SSD di Geografia) del 20%, e una di valutazioni Limitate poco al di sopra del 10% contro una media appena al di sotto del 12%. Anche nel caso dell'Antropologia e della Pedagogia oltre il 50% degli articoli classificati Eccellenti sono apparsi su riviste classificate A (comprese le internazionali), anche se queste ultime non registrano percentuali di valutazioni Eccellenti significativamente più alte della media.

Emerge inoltre che, in tutti i settori, percentuali superiori al 40 e spesso al 50% di tutti i prodotti presentati alla VQR nella categoria "articolo su rivista" sono stati pubblicati su riviste di fascia A (per questo dato e per i precedenti cfr. Tabelle 2.33-2.38). E' possibile che questo dato sia stato in qualche misura influenzato dalle classificazioni effettuate *ex ante* dal GEV; nondimeno esso riflette anche un giudizio ampiamente condiviso dalle comunità scientifiche di riferimento circa la qualità e la reputazione delle riviste in questione.

3 La valutazione di area delle strutture

Nel presentare i risultati più significativi del proprio lavoro, il GEV 11 ritiene di dover attirare l'attenzione del lettore sul fatto che le classifiche di strutture e dipartimenti, della cui solidità il GEV è convinto, vanno però lette e interpretate ricordando il modo in cui la VQR stessa è stata concepita, e in particolare il peso assegnato alla valutazione dei prodotti presentati (Eccellenti=1; Buoni=0,8; Accettabili=0,5; Limitati=0; Prodotti mancanti = -0,5). Al contrario del precedente esercizio di valutazione (la Valutazione Triennale della Ricerca 2000-03), che aveva tracciato una mappa della sola qualità, trascurando il valore complessivo di Atenei (le cui graduatorie erano appunto basate su un numero molto minore di prodotti, scelti dagli Atenei stessi in base alla loro qualità) e dipartimenti, la VQR disegna perciò una mappa della qualità media delle strutture valutate. Specie la posizione dei grandi e "vecchi" Atenei, gonfiati dall'*ope*



legis del 1980, potrebbe perciò soffrire a causa della presenza al loro interno di sacche significative di docenti e ricercatori poco produttivi capaci di oscurare quella di punte di eccellenza anche importanti. Più in generale, c'è la possibilità che—dato il legame tra produttività scientifica e età anagrafica, particolarmente forte in molti SSD—gli Atenei con organici “anziani” risultino peggiori di Atenei più giovani, che godono di una rendita di posizione destinata fatalmente ad esaurirsi.

Al fine di tenere conto, nei limiti del possibile, anche di questi elementi vengono perciò presentati dati sulla distribuzione per struttura dei prodotti Eccellenti per ciascun SSD (cfr. Tabelle 3.1-3.34) e sulle percentuali di soggetti parzialmente o del tutto inattivi per struttura in ciascuna sub-area (cfr. Tabelle 3.35-3.36).

Seguono quindi i risultati di sintesi della valutazione dell'area 11. Le Tabelle riportano la graduatoria delle Università ed Enti di Ricerca ordinati in base al voto medio e all'indicatore IRAS1 ottenuti dai prodotti della ricerca conferiti per la valutazione. In particolare, il voto medio (indicatore *I*) è il rapporto fra la somma dei punteggi ottenuti dai singoli prodotti appartenenti alla struttura e il numero di prodotti attesi per la stessa. Inoltre, IRAS1(%) indica il rapporto tra il punteggio complessivo raggiunto da una struttura in una data area e la valutazione complessiva dell'area stessa.

I risultati di tale analisi comparativa sono riferibili ai prodotti conferiti da più di 3 soggetti valutati (> 9 prodotti per le Università e >18 per gli Enti di Ricerca e Consorzi). Sulla base del Bando, ai singoli prodotti presentati vengono assegnati pesi 1, 0.8, 0.5 e 0a seconda che siano valutati Eccellenti, Buoni, Accettabili o Limitati; ai prodotti mancanti è assegnato peso - 0.5, ai non valutabili è assegnato peso -1, e in casi accertati di plagio o frode si ha peso -2.

Indicando rispettivamente con $n_{i,j}E$, $n_{i,j}B$, $n_{i,j}A$, $n_{i,j}L$, $n_{i,j}MIS$, $n_{i,j}NV$ e $n_{i,j}PL$ il numero di prodotti Eccellenti, Buoni, Accettabili, Limitati, Mancanti, Non Valutabili e Plagio della struttura i -esima nell'Area scientifico-disciplinare j -esima, si ottiene la valutazione complessiva $v_{i,j}$ della struttura i -esima nell'Area j -esima come:

$$v_{i,j} = n_{i,j}E + 0.8 * n_{i,j}B + 0.5 * n_{i,j}A + 0 * n_{i,j}L - 0.5 * n_{i,j}MIS - n_{i,j}NV - 2 * n_{i,j}PL$$



Nelle due sezioni seguenti (3.1 e 3.2) proporremo due indicatori di qualità della ricerca indipendenti dalla numerosità dei soggetti valutati della struttura afferenti all'Area (sezione 3.1) e l'indicatore $IRAS1_{i,j}$ definito nel decreto ministeriale e nel bando VQR, che tiene conto invece sia della qualità della ricerca che della numerosità dei soggetti valutati della struttura afferenti all'Area (sezione 3.2).

I due indicatori proposti nella sezione 3.1 e l'indicatore $IRAS1_{i,j}$ del Bando si riferiscono a una struttura e a un'Area particolare, e non sono quindi affetti da eventuali difformità nelle valutazioni inter-Area.

3.1 Gli indicatori di qualità delle strutture nell'area

3.1.1 Premessa

Gli indicatori proposti sono indicatori di qualità indipendenti dalla numerosità dei soggetti valutati dell'Area nella struttura. Non tenendo conto delle dimensioni della struttura stessa, non possono essere utilizzati da soli per la distribuzione delle risorse, ma devono essere integrati (o sostituiti *in toto*) dall'indicatore $IRAS1_{i,j}$, che tiene conto sia della qualità della ricerca che delle dimensioni della struttura nell'Area. I due indicatori forniscono però informazioni utili sulla qualità della ricerca delle strutture in una determinata Area scientifica.

Prima di presentare gli indicatori e i risultati dell'intero processo valutativo è doveroso premettere che, poiché la VQR non può diffondere i risultati delle singole valutazioni per ragioni di privacy, in tutte le tabelle che seguono sono escluse tutte le strutture con meno di 10 prodotti attesi, ossia:

Consorzio Interuniversitario di Neuroscienze denominato "Istituto Nazionale di Neuroscienze"

Consorzio Nazionale Interuniversitario per le Scienze del Mare – CONISMA

Libera Università Internazionale di Studi Sociali "Guido Carli" LUISS-Roma



Scuola Internazionale Superiore di Studi Avanzati di TRIESTE

SUM - Istituto Italiano di SCIENZE UMANE di FIRENZE

UNICUSANO Università degli Studi Niccolò Cusano -Telematica Roma

Università Commerciale "Luigi Bocconi" MILANO

Università degli Studi "Magna Graecia" di CATANZARO

Università degli Studi "Mediterranea" di REGGIO CALABRIA

Università degli Studi del SANNIO di BENEVENTO

Università degli Studi di CAMERINO

Università Politecnica delle MARCHE

Università Telematica GUGLIELMO MARCONI

Università Telematica PEGASO

3.1.2 Il primo indicatore

Indicando con $n_{i,11}$ il numero di prodotti attesi per la VQR della struttura i -esima nell'area 11, e con $v_{i,11}$ il suo peso, o "voto", ottenuto in base alla valutazione di classe, il **primo indicatore** $I_{i,11}$ è definito da:

$$I_{i,11} = \frac{v_{i,11}}{n_{i,11}} \quad (1)$$

3.1.3 Il secondo indicatore

Il **secondo indicatore** è $R_{i,11}$ dato da:

$$R_{i,11} = \frac{\frac{v_{i,11}}{n_{i,11}}}{\frac{\sum_{i=1}^{N_{ST}} v_{i,11}}{N_{11}}} = \frac{I_{i,11}}{V_{11}/N_{11}} \quad (3)$$



dove V_I e N_{11} indicano la valutazione complessiva e il numero totale di prodotti attesi nell'area 11, vale a dire:

$$V_{11} = \sum_{i=1}^{N_{ST}} v_{i,11} \quad , \quad N_{11} = \sum_{i=1}^{N_{ST}} n_{i,11} \quad (4)$$

L'indicatore $R_{i,11}$ rappresenta il rapporto tra il voto medio ricevuto dai prodotti della struttura i -esima nell'area 11 e il voto medio ricevuto da tutti i prodotti dell'area 11. Esso consente una misura diretta della qualità relativa della ricerca in una certa area espressa da una determinata struttura: valori inferiori a 1 indicano una produzione scientifica di qualità inferiore alla media di area, valori superiori a 1 indicano una qualità superiore alla media.

Il **secondo indicatore** $S_{i,j}$, anch'esso compreso tra zero e 1, è definito come

$$S_{i,j} = \frac{\frac{v_{i,j}}{n_{i,j}}}{\sum_{i=1}^{N_{ST}} \frac{v_{i,j}}{n_{i,j}}} = \frac{A_{i,j}}{\sum_{i=1}^{N_{ST}} A_{i,j}} \quad , \quad \sum_{i=1}^{N_{ST}} S_{i,j} = 1 \quad (2)$$

dove N_{ST} è il numero totale di strutture.

L'indicatore $S_{i,j}$ è una versione normalizzata del primo indicatore $I_{i,j}$, e può essere utilizzato per una parziale ripartizione delle risorse in un'area fra le strutture che tenga conto soltanto della qualità e non delle dimensioni delle strutture stesse.

3.1.4 Ranking delle strutture sulla base dei due indicatori

Le Tabelle 3.37-3.48 indicano rispettivamente le graduatorie di Università e degli Enti di Ricerca sulla base dei due indicatori I ed R . Per la sub-area non bibliometrica le strutture sono suddivise in grandi, medie e piccole in base al numero di prodotti attesi: G=grandi ($n \geq 300$), M=medie ($299 \geq n \geq 100$), P=piccole ($n < 100$). Invece, per la sub-area non bibliometrica e a livello di subGEV (per Storia, Filosofia, i settori non bibliometrici di Pedagogia e Psicologia) le strutture sono state suddivise in grandi, medie e piccole sempre in base al numero di prodotti attesi, ma con soglie differenti: G=grandi ($n \geq 100$), M=medie ($99 \geq n \geq 30$), P=piccole ($n < 30$).



In queste tabelle, inoltre, viene fornita (alla voce *Posizione in graduatoria*) anche l'indicazione della posizione che ciascuna struttura ricoprirebbe in una graduatoria basata sui due indicatori I ed R che, però, non tenesse conto delle diverse dimensioni delle strutture. Per i settori di Discipline Demoetnoantropologiche, Geografia, Scienze del Libro e del Documento, nonché per i settori bibliometrici di Pedagogia, le strutture non sono state suddivise in grandi, medie e piccole. Inoltre è mostrata la distribuzione percentuale dei prodotti della ricerca nelle 4 classi finali di merito e la percentuale dei prodotti penalizzati. Le Tabelle non includono le università con meno di 10 prodotti attesi e gli Enti e i Consorzi con meno di 19 prodotti attesi.

Le Tabelle 3.49-3.78 riportano invece le graduatorie delle Università ordinate in base al voto medio ottenuto dai prodotti della ricerca conferiti relativamente a ciascun SSD dell'area 11. Nelle tabelle non sono incluse le Università con meno di 10 prodotti attesi e gli Enti di Ricerca e Consorzi con meno di 19 prodotti attesi; l'assenza delle graduatorie per alcuni piccoli SSD (come M-FIL/07 e M-FIL/08) è appunto legata al fatto che non vi sono Università con più di 10 prodotti attesi in quegli SSD.

3.2 L'indicatore $IRAS1_{i,j}$ del Bando VQR

L'indicatore $IRAS1_{i,j}$ è definito nel Bando VQR come rapporto tra il punteggio complessivo raggiunto da una struttura in una data area e la valutazione complessiva dell'Area stessa:

$$IRAS1_{i,j} = \frac{v_{i,j}}{\sum_{i=1}^{NST} v_{i,j}} = \frac{v_{i,j}}{V_j} \quad (5)$$

Esso può essere scritto come il prodotto di un indicatore di qualità relativa dei prodotti presentati da un certa struttura in una data area per un indicatore della dimensione della struttura nella stessa area. L'indicatore di qualità è dato dal rapporto tra il voto medio ricevuto dai prodotti della struttura i -esima nell'area j -esima rispetto al voto medio ricevuto da tutti i prodotti dell'area j -esima, e corrisponde al terzo indicatore $R_{i,j}$ definito nella (3), mentre il peso della struttura ($P_{i,j}$) è dato semplicemente dalla quota di prodotti attesi dell'area j -esima dovuti alla struttura i -esima:



$$IRAS1_{i,j} = \frac{\frac{v_{i,j}}{n_{i,j}}}{\frac{\sum_{i=1}^n v_{ij}}{N_j}} \times \frac{n_{i,j}}{N_j} = \frac{A_{i,j}}{V_j / N_j} \times \frac{n_{i,j}}{N_j} = R_{i,j} \times P_{i,j} \quad (6)$$

L'indicatore $IRAS1_{i,j}$ ridefinisce il peso di una struttura in un'area, misurato dalla quota dei prodotti attesi, sulla base della qualità relativa dei prodotti attesi stessi. Come tale, $IRAS1$ è un indicatore utile soprattutto per la ripartizione dei fondi tra strutture in una medesima area, in quando tiene conto insieme della qualità e del peso relativo di una struttura.

Le tabelle 3.79-3.81 mostrano il ranking delle strutture sulla base dell'indicatore $IRAS1$, per le due sub-aree, distinguendo tra Università (divise in grandi, medie e piccole strutture come nelle Tabelle ordinate sulla base dell'indicatore I) e Centri di Ricerca. Poiché la graduatoria dei Centri di Ricerca della sub-area non bibliometrica per $IRAS1$ coincide con quella alla Tabella 3.39, essa non è stata duplicata. Anche in queste tabelle, inoltre, viene fornita (alla voce *Posizione in graduatoria*) l'indicazione della posizione che ciascuna struttura ricoprirebbe in una graduatoria basata sull'indicatore $IRAS1$ e che però non tenesse conto delle diverse dimensioni delle strutture.

3.3 Commenti sul significato degli indicatori di struttura di area

Il primo indicatore, $I_{i,j}$, è un indice di qualità della produzione scientifica che assume il valore 1 nel caso in cui la struttura abbia presentato tutti i prodotti attesi, e tutti abbiano ottenuto la valutazione di eccellente.

Il secondo indicatore fornisce una indicazione sulla posizione della struttura rispetto alla media di area. Se il suo valore è maggiore di 1, significa che la struttura ha una qualità sopra la media di area, se è minore di 1, sta sotto la media. Anch'esso, opportunamente normalizzato, potrebbe essere utilizzato per una distribuzione di risorse che premi soltanto la qualità indipendentemente dalle dimensioni delle strutture.

Infine, l'indicatore $IRAS1_{i,j}$, definito dal decreto ministeriale e dal Bando, incorpora la valutazione puramente qualitativa con le dimensioni della struttura, e può essere utilizzato per una distribuzione delle risorse che sia vista come una alterazione, basata sulla qualità, della distribuzione puramente proporzionale. Infatti, se in tutte le strutture i prodotti ottenessero la



stessa valutazione media, l'indicatore rifletterebbe soltanto il numero relativo di prodotti presentati, e quindi, il peso della struttura nella specifica area valutata.

4 La valutazione di area dei dipartimenti

4.1 Premessa

La VQR ha, tra i suoi compiti, quello di fornire alle strutture un *ranking* dei dipartimenti che possa essere utilizzato come informazione dagli organi decisionali delle strutture nella distribuzione interna delle risorse.

Gli statuti degli atenei approvati a valle della Legge 240 presentano diverse tipologie di dipartimenti. Le più frequenti sono rappresentate da:

1. Dipartimenti che inglobano in toto dipartimenti più piccoli preesistenti;
2. Dipartimenti che raccolgono in varia misura frange di dipartimenti preesistenti, con una struttura fortemente composita e difficilmente ascrivibile a una (o due) aree VQR.

In entrambi i casi, occorre costruire degli indicatori di dipartimento a partire dalle valutazioni dei prodotti associati ai soggetti valutati di quel dipartimento. Anche in questo caso, come già fatto per le strutture, è importante far sì che il risultato finale non sia influenzato da difformità di valutazione inter-Area. Inoltre, poiché la VQR non può diffondere i risultati delle valutazioni relative a singoli individui o ad essi riconducibili, è opportuno che le tabelle prendano in considerazione solo le strutture con almeno 10 lavori attesi (e quindi conferiti da almeno 3 soggetti valutati) per Area (o SSD nel caso di classifiche per SSD) riducendo così la possibilità di risalire alle valutazioni di singoli studiosi. Nel caso degli enti di ricerca, dov'erano attesi 6 prodotti per persona, esse prendono in considerazione solo le strutture con almeno 19 lavori attesi (e quindi conferiti da almeno 3 soggetti valutati) per Area (o SSD nel caso di classifiche per SSD) riducendo così la possibilità di risalire alle valutazioni di singoli studiosi.

Indicando rispettivamente con $n_{i,j,k}E$, $n_{i,j,k}B$, $n_{i,j,k}A$, $n_{i,j,k}L$, $n_{i,j,k}MIS$, $n_{i,j,k}NV$ e $n_{i,j,k}PL$ il numero di prodotti Eccellenti, Buoni, Accettabili, Limitati, Mancanti, Non Valutabili e Plagio del dipartimento k -esimo della struttura i -esima nell'Area scientifico-disciplinare j -esima, si ottiene la valutazione complessiva $v_{i,j,k}$ del dipartimento k -esimo della struttura i -esima nell'Area j -esima come:

$$v_{i,j,k} = n_{i,j,k}E + 0.8*n_{i,j,k}B + 0.5*n_{i,j,k}A + 0*n_{i,j,k}L - 0.5*n_{i,j,k}MIS - n_{i,j,k}NV - 2*n_{i,j,k}PL .$$



4.2 Gli indicatori di qualità di area del dipartimento

In questa sezione introdurremo due indicatori di qualità di area dei dipartimenti (I ed R) indipendenti dalla numerosità dei soggetti valutati dell'area nei dipartimenti. Non tenendo conto delle dimensioni dei dipartimenti, non possono essere utilizzati da soli per la distribuzione delle risorse, ma devono essere integrati (o sostituiti in toto) dall'indicatore $IRDI_{i,j,k}$, che tiene conto sia della qualità della ricerca che delle dimensioni del dipartimento nell'area. I due indicatori forniscono però informazioni utili sulla qualità della ricerca del dipartimento in una determinata area scientifica.

4.2.1 Il primo indicatore

Indicando con $n_{i,11,k}$ il numero di prodotti attesi per la VQR del dipartimento k -esimo della struttura i -esima nell'area 11, il **primo indicatore** $I_{i,11,k}$ è dato da:

$$I_{i,11,k} = \frac{v_{i,11,k}}{n_{i,11,k}}$$

Questo indicatore rappresenta quindi il voto medio ottenuto dal dipartimento k -esimo della struttura i -esima nell'area 11.

4.2.2 Il secondo indicatore

Il **secondo indicatore** è dato da

$$R_{i,11,k} = \frac{\frac{v_{i,11,k}}{n_{i,11,k}}}{\frac{\sum_{i=1}^{NST} v_{i,11}}{N_{11}}} = \frac{I_{i,11,k}}{V_{11}/N_{11}} \quad (9)$$

dove V_{11} e N_{11} indicano la valutazione complessiva e il numero totale di prodotti attesi nell'area 11.

L'indicatore $R_{i,11,k}$ rappresenta il rapporto tra il voto medio ricevuto dai prodotti del dipartimento k -esimo della struttura i -esima nell'area 11 e il voto medio ricevuto da tutti i prodotti dell'area 11. Esso consente una misura diretta della qualità relativa della ricerca in una



certa area espressa da un determinato dipartimento: valori inferiori a 1 indicano una produzione scientifica di qualità inferiore alla media, valori superiori a 1 indicano una qualità superiore alla media.

4.2.3 *Graduatorie dei Dipartimenti sulla base dei due indicatori*

Nelle tabelle 4.1-4.12 le graduatorie dei Dipartimenti delle Università ante e post L240 sulla base dei due indicatori (*I* ed *R*), per sub-area e per subGEV o sottoinsiemi omogenei degli stessi. Le graduatorie sono suddivise sulla base delle dimensioni dei dipartimenti: a livello di sub-area tra grandi (con almeno 100 prodotti), medi (da 30 a 99 prodotti) e piccoli (non più di 29 prodotti); a livello di subGEV tra grandi (con almeno 30 prodotti) e piccoli (non più di 29 prodotti). La suddivisione non è stata applicata alle graduatorie fornite per Discipline Demoetnoantropologiche, Geografia, Scienze del Libro e del Documento e per i settori bibliometrici di Pedagogia. Le tabelle 4.13-4.41 forniscono invece le graduatorie dei Dipartimenti delle Università post L240 per SSD, ordinate per voto medio e distribuzione dei prodotti nelle classi di merito. Le tabelle 4.50-4.54 forniscono infine le graduatorie delle sottostrutture degli Enti di ricerca, per sub-area, ordinate per voto medio e distribuzione dei prodotti nelle classi di merito: esse forniscono anche indicazioni relative al valore dell'indicatore IRD1 (sul quale v. appresso).

Nelle tabelle che riguardano i dipartimenti successivi alla Legge 240 si trova in alcuni casi un dipartimento indicato con *n.d.* (non definito). A quel dipartimento sono stati attribuiti i soggetti valutati accreditati dalle strutture ma non assegnati ad alcun dipartimento post-240. I motivi possono essere che:

- I soggetti valutati erano in servizio alla data del Bando (7 novembre 2011) ma hanno cessato la loro attività prima dell'attribuzione del personale ai nuovi dipartimenti post-240;
- Le università non avevano ancora completato la costituzione dei nuovi dipartimenti alla scadenza fissata per l'attribuzione dei soggetti valutati (scadenza che ha subito varie proroghe fino alla primavera 2013)



4.3 L'indicatore $IRD1_{i,j,k}$ del Bando VQR

L'indicatore $IRD1_{i,j,k}$ è definito nel Bando VQR come rapporto tra il punteggio complessivo raggiunto da un dipartimento in una data area rispetto alla valutazione complessiva dell'Area stessa:

$$IRD1_{i,j,k} = \frac{v_{i,j,k}}{\sum_{i=1}^{N_{ST}} v_{i,j}} \quad (10)$$

Esso può essere scritto come il prodotto di un indicatore di qualità relativa dei prodotti presentati da un certo dipartimento in una data area per un indicatore delle dimensioni del dipartimento nella stessa area. L'indicatore di qualità è dato dal rapporto tra il voto medio ricevuto dai prodotti del dipartimento k -esimo della struttura i -esima nell'area j -esima e il voto medio ricevuto da tutti i prodotti dell'area j -esima, e corrisponde al terzo indicatore $R_{i,j,k}$ definito nella (9), mentre la dimensione del dipartimento ($P_{i,j,k}$) è data semplicemente dalla quota di prodotti dell'area j -esima dovuti al dipartimento k -esimo della struttura i -esima:

$$IRD1_{i,j,k} = \frac{\frac{v_{i,j,k}}{n_{i,j,k}}}{\frac{\sum_{i=1}^n v_{i,j}}{N_j}} \times \frac{n_{i,j,k}}{N_j} = R_{i,j,k} \times P_{i,j,k} \quad (11)$$

L'indicatore $IRD1_{i,j,k}$ è quindi un indicatore che ridefinisce il peso di un certo dipartimento di una certa struttura in una certa area, misurato dalla quota dei prodotti attesi, sulla base della qualità relativa dei prodotti attesi stessi. Come tale, $IRD1$ è un indicatore utile soprattutto per la ripartizione dei fondi tradipartimenti della stessa struttura in una medesima area, in quanto tiene conto insieme della qualità della ricerca e del peso relativo del dipartimento.

Le graduatorie *all'interno delle strutture* dei Dipartimenti delle Università, ante e post L240 sono fornite nelle tabelle 4.42-4.49, sia sulla base di $IRD1$ che su quella del voto medio. Poiché l'indicatore $IRD1$ (vedi formula (11)) è il prodotto di un indicatore di qualità (R) e di dimensione (P), nelle tabelle si riporta sia il valore di $IRD1$ che quello di P . Evidentemente, i dipartimenti con $IRD1 > P$ sarebbero premiati dalla valutazione rispetto a una distribuzione di risorse puramente proporzionale al numero di prodotti attesi.



5 Analisi dei risultati

Un'analisi puntuale dei dati della VQR non è compito del GEV che li ha prodotti, che è tenuto piuttosto a fornire gli strumenti critici necessari alla loro corretta lettura, e quindi a spiegare—come abbiamo cercato di fare—il modo in cui quei dati sono stati costruiti, i loro punti di forza e quelli di debolezza.

L'estrema ricchezza e varietà dei risultati della VQR, la loro straordinaria capacità di fornire per la prima volta una fotografia precisa e dettagliata della situazione della ricerca nelle Università italiane, scendendo fino al livello dei dipartimento, e di concentrazioni significative di SSD al loro interno, richiederà quindi analisi sofisticate, che sperabilmente alimenteranno politiche mirate e razionali. E' comunque evidente che questi risultati possono e debbono essere letti a molti livelli.

Il primo è naturalmente quello rappresentato dalla distribuzione qualitativa degli Atenei e degli enti di ricerca italiani che almeno in Area 11 risulta fortemente piramidale, con molti atenei—specie se grandi e purtroppo concentrati, ma non esclusivamente, nelle regioni centro-meridionali—appesantiti da un'area di prodotti limitati (con punteggio eguale a zero) e da una non trascurabile presenza di prodotti mancanti (vale a dire di docenti e ricercatori parzialmente o totalmente inattivi). Al contrario della VTR, che si limitava a valutare i prodotti migliori di ciascun ateneo, e quindi restituiva un'immagine meno stratificata, la VQR ha preso infatti in esame un campione della produzione “migliore” di tutti i docenti e i ricercatori italiani, facendo così emergere quelle aree e spesso appesantendo la posizione di realtà dove pure non mancano isole e reti di lavoro buono o eccellente. E' qui il caso di ricordare come la percentuale di lavori attesi ma non presentati alla VQR è stata pari a zero in alcuni settori (come M-EDF/02, M-FIL/07, M-PSI/02 e 03, M-STO/08) mentre ha raggiunto e superato il 5%, ovverosia il doppio di una media di area pari al 2,5%, in altri (M-PSI/08, M-STO/06, 07 e 09).

Il secondo livello è quello, molto interessante, della distribuzione qualitativa dei dipartimenti, che si dispone anch'essa, e per le stesse ragioni, su una piramide molto ripida. I dati raccolti permettono inoltre di produrre una classifica dei dipartimenti pre-Legge 240 e una dei dipartimenti post Legge 240. Nel caso di dipartimenti che abbiano effettuato politiche di “acquisto” e accorpamento, ovverosia di “dismissione” di loro pezzi, sarà quindi possibile avere a disposizione materiale per un'analisi e una valutazione delle politiche seguite sia dagli atenei, che dai gruppi dirigenti dei dipartimenti interessati.



Altrettanto importanti sono i dati relativi alla distribuzione qualitativa dei SSD sul territorio nazionale, e nei dipartimenti, che per la prima volta forniscono una mappa precisa della situazione al loro interno, e rendono possibile una comparazione indiretta (che passa cioè attraverso il confronto della posizione occupata dagli SSD di ciascun Ateneo o ente di ricerca nella graduatoria nazionale dei rispettivi SSD) tra nuclei di SSD diversi.

Importanti sono anche i dati relativi alla distribuzione delle eccellenze, calcolate come concentrazione per Ateneo o ente di ricerca delle % di E di ciascun SSD, che indicano con precisione gli Atenei e i dipartimenti da preferire per lo studio di determinate discipline e agli organi di governo su cosa puntare.

La relativa omogeneità dei risultati della valutazione nelle Aree e nei macro-settori bibliometrici (indicata per esempio dalla relativa stabilità della % di E e E+B nelle diverse aree e nel caso dell'Area 11 tra i 10 SSD bibliometrici) contrasta con la già ricordata grande variabilità presentata dai risultati dei macro-settori e dei SSD *peer* che—come già ricordato nella *Introduzione*—presentano:

1. Risultati E (e E+B) significativamente e strutturalmente inferiori, e risultati L altrettanto strutturalmente superiori, a quelli dei macro-settori bibliometrici;
2. Una forte variabilità nella distribuzione E-B-A-L tra SSD diversi.

I risultati della valutazione dei prodotti della Psicologia mostrano per esempio una collocazione nelle quattro fasce di merito che vede un valore superiore al 35% nella fascia E, un valore superiore all'11% nella fascia B, un valore superiore al 24% nella fascia A e un valore poco superiore al 26% nella fascia L. Le E dei SSD *peer* superano invece il 30% solo in alcuni piccoli SSD (con una punta del 44% per M-FIL/08) e oscillano tra il 6 e il 21% in quelli più grandi (cioè con oltre 500 prodotti attesi); notevole appare la variabilità dei loro risultati, anche se si fanno confronti tra SSD di dimensioni analoghe (si va per esempio dal 6% di M-GGR/01 al 19% circa di M-FIL/03 e M-STO/01, tutti settori con circa 500 prodotti attesi, e dal 13% di M-PED/01 al 17% di M-STO/02, per quanto riguarda settori con circa 800 prodotti attesi), a conferma che la VQR *peer* ha dato risultati credibili, di grande valore e interesse, soprattutto a livello di SSD.

Restando ai macro-settori *peer*, che rappresentano la maggioranza dell'Area, la VQR ha rivelato un quadro in cui, nelle parole di un EV, accanto alle punte di eccellenza vi sono vaste aree di dignitosa o modesta operosità, e una consistente minoranza al di sotto dello standard, che ha spesso sottoposto prodotti ai limiti dell'impresentabilità, cosa tanto più grave se si considera



che si trattava di segnalare le proprie tre opere migliori su un arco di sette anni. Al di sotto di questa minoranza vi è il già ricordato, consistente spezzone che non è stato nemmeno in grado di presentare, in tutto o in parte, i tre prodotti richiesti.

Vi è quindi nell'Università e negli enti di ricerca italiani una non trascurabile percentuale di studiosi poco e malamente produttivi, o addirittura improduttivi, che non si sono fin qui nemmeno posti il problema di avere una sia pur solo dignitosa produzione scientifica. La VQR la ha fatta emergere con chiarezza, e starà ora agli Atenei e agli enti di ricerca trovare il modo di stimolarla a riqualificare i suoi comportamenti e soprattutto di impedirne la riproduzione.

In generale, il punto di maggior sofferenza della produzione scientifica dell'Area 11, di quella dignitosa ma anche, e non raramente, di quella di eccellenza, è rappresentato dalla sua scarsa internazionalizzazione, che non è tanto o solo una questione di lingua, ma spesso anche di mancato rapporto col dibattito internazionale attorno ai temi trattati dalle singole ricerche.

Il livello di internazionalizzazione varia tra l'altro notevolmente da SSD a SSD e spesso gli SSD più internazionalizzati sono paradossalmente quelli che hanno sofferto di più nella VQR, anche perché la loro parte già internazionalizzata ha usato con rigore la domanda sull'internazionalizzazione, che è stata invece interpretata in modo molto estensivo dai SSD poco internazionalizzati.

Come ha scritto un membro del GEV, in SSD come quelli di storia “molti *referee* si sono trovati in difficoltà a valutare l'internazionalizzazione, che da sola insiste per ben 1/3 della valutazione. Ottime ricerche di storia hanno base archivistico-documentaria, o filologico-testuale, e riguardano contesti specifici, per lo più nazionali o regionali, come accade nella gran parte della ricerca storica, non solo in Italia. In questi casi, l'internazionalizzazione è per definizione carente, e al più può esser rilevata solo nei riferimenti comparativi, fatti salvi, naturalmente, i rari studi su contesti non italiani (ma non perciò necessariamente ‘internazionali’), o quelli scritti in lingua estera e pubblicati in riviste o con editori non italiani. Per contro, la scheda non offre il modo di rilevare il tasso di nazionalizzazione, ovvero la circolazione in ambito nazionale e non solo locale”. Anche in altri SSD è stato notato un “uso pressoché esclusivo dell'italiano”, che in qualche caso però “rispecchia gli usi tradizionali” di settori “fortemente ancorati alla tutela del plurilinguismo nelle sedi di confronto internazionale” e nei quali la spinta verso un uso più ampio dell'inglese si è cominciata ad affermare solo negli ultimi anni. La valutazione dell'internazionalizzazione, di sicuro fondamentale anche per il futuro della ricerca italiana nelle scienze umane, è stata così anche fonte di discrepanza e



disomogeneità di giudizio a seconda delle decisioni del *referee* di adottare un criterio rigido e restrittivo, oppure elastico e estensivo, della categoria.

La cosa positiva è che di questa scarsa internazionalizzazione la VQR ci rimanda oggi un'immagine precisa. Questa immagine può essere da sola, e tanto più in presenza di opportuni interventi, un forte stimolo al superamento di questo *handicap*, che pare tanto più grave nei casi in cui la qualità della produzione scientifica italiana è tale da permetterle di giocare un ruolo di primo piano nella ricerca internazionale.

Nelle parole di un altro componente del GEV, l'analisi di massa della produzione scientifica su sette anni ha fatto emergere in taluni SSD *peer* anche una scarsa "presenza di sedi adeguate a valutare, anche informalmente, gli standard qualitativi e la direzione della ricerca disciplinare (riviste, congressi, ecc.)". L'impressione è che in alcuni SSD coesistano aree di ricerca isolate le une dalle altre, tanto da configurare una situazione in cui membri dello stesso SSD non sanno bene cosa fa la *loro* comunità nazionale e non dialogano con essa, anche a causa "dell'assenza di riviste chiave per la progressione della carriera e per la certificazione dei vari paradigmi", un'assenza cui forse la VQR stessa ha cominciato a porre rimedio.

Di grande interesse sono anche i dati relativi alla distribuzione nelle classi E-L dei prodotti presentati dalle tre fasce di docenza (o di ricerca per quel che riguarda gli enti), certo limitatamente ai prodotti di autore singolo che costituiscono tuttavia—come si è detto—la stragrande maggioranza dei prodotti della parte non bibliometrica dell'Area 11. Come già detto in precedenza, nel caso di lavori con più di un autore non è facile distinguere il contributo di un singolo autore, specie laddove gli autori siano molti. Si è perciò supposto che ciascuno degli autori abbia contribuito nella stessa misura alla stesura del lavoro, attribuendo quindi la "qualità media dei lavori presentati" (in genere tre) alla "qualità del soggetto valutato". In ogni caso, le statistiche vanno interpretate con cautela, e servono a evidenziare le tendenze che emergono tra gruppi consistenti di ricercatori (strutture, dipartimenti, SSD, ecc.) e non per analisi o valutazioni di casi singoli. Ciò vale non solo per la distribuzione delle classi dei prodotti per fascia di docenza, ma anche per le altre informazioni relative alla valutazione di prodotti altrimenti riferiti a soggetti valutati (cfr. Tabelle 5.1-5.6).

Sia pure con variazioni significative a livello di SSD, risulta in particolare che gli ordinari hanno nel loro complesso valutazioni superiori a quelle dei ricercatori, che ricevono—con le eccezioni di cui si è detto—valutazioni superiori a quelle degli associati, forse a significare un miglioramento nella qualità delle assunzioni di giovani studiosi nel primo decennio del nuovo



millennio. L'età media degli studiosi nei singoli Atenei, e la sua influenza sui risultati della VQR, meriterebbe quindi attenzione particolare.

Preoccupante è infine il dato relativo alla distribuzione delle classi per macroregione geografica (cfr. Tabelle 5.7-5.16), da cui risulta che—sebbene anche in questo con significative eccezioni a livello di SSD—gli Atenei del Meridione presentano una percentuale minore di prodotti eccellenti rispetto a quelli del Centro, del Nord-Ovest e del Nord-Est, e una maggiore di prodotti limitati. Le variazioni delle classi intermedie sono invece meno significative, nel mentre il Centro, pur ben piazzato per quanto riguarda i prodotti eccellenti è anch'esso appesantito da una percentuale maggiore di prodotti limitati.

Dei problemi posti dalla selezione dei *referee* e poi dal loro comportamento si è già detto. Qui torniamo a sottolineare che questa è sicuramente una delle fonti della forte variabilità dello strumento *peer*, a sua volta legata alla sua intrinseca soggettività. Quest'ultima potrebbe probabilmente essere meglio controllata in futuro attraverso il perfezionamento delle schede di valutazione. Abbiamo già affrontato i problemi posti dalla sezione dedicata all'internazionalizzazione, e detto di quelli sollevati dalla “assenza di un metro di valutazione riguardante la consistenza dei prodotti” in SSD in cui la monografia tende ancora a prevalere sugli articoli e dove quindi l'adozione di “una stessa griglia di valutazione per prodotti evidentemente non assimilabili (ad es. la voce di un dizionario biografico di poche pagine e una monografia di oltre 600), che hanno e devono avere un peso molto diverso” ha spesso messo in difficoltà i *referee*, costringendoli a scelte che non sempre sono state univoche.

Tuttavia, in un campo, come quello umanistico, caratterizzato da un progresso meno lineare e cumulativo del sapere, anche “l'incidenza attribuita alle capacità del prodotto di ‘stimolare nuove ricerche’ o di ‘far avanzare le conoscenze’ (valutazioni assai simili, ma riferite nella scheda a sezioni diverse)” ha posto i *referee* di fronte a scelte difficili, cui era possibile dare risposte diverse.

Nell'introduzione si è già accennato al problema specifico posto dai SSD più piccoli, che presentano percentuali di E talvolta quasi doppie (in un caso più che doppie) rispetto alla media dei SSD non bibliometrici, fino a 5-6 volte superiori a quelle dei SSD più grandi, e spesso pari o superiori a quelle riportate dai SSD valutati in bibliometria. Si tratta evidentemente di un problema sistemico, che può essere ricondotto alla “vicinanza” (*collaboration distance*), molto difficile da eludere in SSD di limitate dimensioni e ad alta specializzazione, dove tutti si conoscono e parlano linguaggi comuni, spesso formalmente rigorosi. Questi risultati atipici non



sono insomma segno di “collusione” ma frutto naturale di caratteristiche specifiche che la *peer review* non è in grado di contrastare. Essi inoltre non confliggono necessariamente con la qualità di questi SSD, semmai—ma non per colpa—afflitti da un’inevitabile autoreferenzialità che nei casi migliori è difesa di standard tradizionalmente elevati di competenza, spesso condivisi da piccole comunità anche fortemente internazionalizzate, in quelli peggiori può trasformarsi in un’autarchia confinante con il conservatorismo, e pone comunque sempre il problema—di non facile soluzione—di un suo superamento.

Riassumendo, la *peer review* ha quindi permesso una mappatura accurata e credibile della qualità e della sua distribuzione *all’interno* degli SSD, con la relativa eccezione di quelli molto piccoli. Ma ha dimostrato di non essere in grado di agire da unità di misura generale a causa della grande variabilità dei suoi risultati, a sua volta legata al suo soggettivismo, alle diverse culture della valutazione prevalenti nei vari SSD e tra i *referee* ad essi appartenenti, alle dimensioni di questi ultimi, al loro atteggiamento e alla loro situazione rispetto all’internazionalizzazione, ecc. “Alla luce della conoscenza che ciascuno di noi ha dei diversi SSD—ha scritto per esempio il coordinatore del subGEV di Filosofia—differenze così forti [come quelle risultanti dalla VQR] appaiono poco plausibili, tanto da rendere poco opportuna la comparazione fra i diversi SSD, o tra gruppi di ricercatori appartenenti a diversi SSD”, mentre quello di Pedagogia ha rimarcato “la presenza di un’eterogeneità intrinseca nei quattro settori MPED...per quanto riguarda la cultura della valutazione”.

Anche la tipologia delle pubblicazioni sottoposte dall’Area 11 alla valutazione rivela la diversità tra i macro-settori e gli SSD bibliometrici, in cui più alta è la percentuale di articoli su rivista (pari al 77% dei prodotti conferiti), spesso indicizzati nelle grandi banche dati internazionali come WoS e Scopus, e quelli *peer*. Qui i prodotti con più autori sono pochissimi, pochi gli articoli su riviste indicizzate (in tutto 556, circa il 25% sul totale degli articoli su rivista, ma poco meno del 6% di tutti i prodotti conferiti nei SSD non bibliometrici) e di fascia A (in tutto essi rappresentano circa il 50% del totale degli articoli su rivista conferiti nei SSD non bibliometrici, e in assoluto poco più del 12% di tutti i prodotti conferiti nei SSD non bibliometrici, e inferiori al previsto le monografie (circa il 33% di tutti i prodotti conferiti nei SSD non bibliometrici), almeno una parte delle quali pubblicate per espliciti fini concorsuali. Emerge invece una forte tendenza a pubblicare contributi in miscelanee, a loro volta spesso *Festschriften*, opere collettive, volumi congressuali o comunque pubblicazioni in cui si scrive su invito, che mette in luce la dipendenza delle pubblicazioni dall’appartenenza a determinate reti, spesso isolate le une dalle altre, piuttosto che a una comunità scientifica nazionale a sua volta parte integrante e integrata di quella internazionale. Come ha notato con accenti autocritici una



collega che si è accorta di aver seguito le stesse pratiche, ciò significa pubblicare in modo dispersivo e “sostanzialmente occasionale, su sollecitazioni abbastanza casuali, in sedi che non sempre consentono una vasta diffusione [a differenza degli articoli, le miscellanee figurano raramente nelle banche dati internazionali e sono quindi di più difficile reperimento, e spesso non vengono recensite dalle buone riviste]. Non ho cioè tenuto conto del fatto che uno dei doveri di chi si impegna scientificamente è non solo quello di produrre personalmente buoni risultati scientifici, ma anche di far progredire la disciplina nel suo insieme, diffondendo il più possibile tali risultati: ciò che si ottiene, oltre che con le monografie, cercando di pubblicare su riviste di prestigio, abitualmente frequentate anche da colleghi stranieri”.

Come nel caso dell'internazionalizzazione e della diffusione di pratiche di valutazione *ex ante* in riviste e case editrici, si può sperare che una delle ricadute positive della VQR sia anche il mutamento, almeno parziale, di queste tradizioni di pubblicazione a favore di nuove che permettano una maggiore integrazione nelle comunità nazionali di riferimento, e di queste in quelle internazionali; siano fondate sulla produzione di opere di più facile reperimento e meglio segnalate; e passino attraverso più rigorosi filtri qualitativi ad opera di quelle stesse comunità.

6 Conclusioni

Nel corso della VQR il GEV 11 si è attenuto al mandato ricevuto. Nonostante problemi la cui individuazione è utile a capire i possibili margini di errore dei dati ottenuti, il GEV conferma la fiducia nei metodi e nelle procedure adottate per il rilevamento, l'elaborazione e la valutazione della ricerca nell'Area 11, e ritiene che:

- a. Tali metodi e tali procedure fossero le migliori tra quelle disponibili;
- b. I risultati ottenuti grazie al loro impiego siano nel complesso solidi ed affidabili, i migliori che fosse possibile conseguire nelle condizioni date.

La VQR è stata insomma una grande impresa, fortemente innovativa. Il GEV 11 auspica perciò che i suoi risultati trovino applicazione nel processo di miglioramento della qualità dell'università e della ricerca italiane, ed è convinto che un loro intelligente utilizzo possa avere ricadute positive importanti, e non solo per chi occupa le posizioni di vertice nei *ranking* da essa prodotti.

Per quel che riguarda il futuro della valutazione della ricerca, in base alla sua esperienza il GEV 11 auspica in particolare che:



- a. La scheda utilizzata nella *peer review* venga riformulata a) tenendo conto, nei SSD dove ciò è opportuno, della diversità dei vari prodotti (riconoscendo per esempio un peso maggiore, a parità di qualità, alle monografie rispetto agli articoli); b) trovando una definizione migliore e più univocamente interpretabile della internazionalizzazione; e c) incoraggiando il breve commento finale di giustificazione dei voti assegnati e rendendo possibile all'EV responsabile di richiederlo in caso di dubbi;
- b. Nei prodotti inviati in bibliometria a) tra quelli con più autori si distingua in base alla posizione dell'autore, rispettando le tradizioni di ciascuna Area o macro-settore (in alcuni casi sarebbe per esempio opportuno ammettere solo i prodotti in cui l'autore sia al primo o all'ultimo posto, oppure, in alternativa, si potrebbe aumentare il numero di prodotti che deve essere presentato da autori che non figurino al primo o all'ultimo posto); b) si ripensi, anche qui nel rispetto delle tradizioni di ciascuna Area o macro-settore, l'uso contemporaneo di due banche dati (ISI e Scopus), che può costituire motivo di squilibrio attribuendo collocazioni più generose sulla base di Scopus rispetto ai prodotti indicizzati ISI; c) si cerchi il modo di "istruire" il sistema a valutare i prodotti in base alla *subject category* ISI per essi più vantaggiosa, che non sempre è risultata essere quella indicata dagli autori;
- c. Vengano ripensati i *rating* delle riviste alla luce dei risultati della VQR.

In generale, il GEV 11 ritiene che:

- a. La VQR abbia fornito elementi utili ai fini di un ripensamento della struttura e del numero delle aree CUN, almeno ai fini della valutazione, suggerendo uno spostamento della tradizionale suddivisione nazionale delle discipline verso il modello europeo (ERC) e comunque verso un maggiore accorpamento di SSD affini da un lato, e una progressiva autonomizzazione di macro-settori dotati di forte personalità, anche a livello internazionale, dall'altro;
- b. L'utilizzo della *peer review*, inevitabile anche a causa della tipologia della produzione scientifica di gran parte degli SSD dell'Area e della scarsa indicizzazione nelle banche dati esistenti dei prodotti presentati, ha sollevato problemi causati dalla sua diversità strutturale rispetto alla bibliometria e dalla grande variabilità dei suoi risultati a seconda della cultura e delle dimensioni di ciascun SSD. Il primo elemento tende a sfavorire le aree valutate con la *peer review* in confronti con quelle



bibliometriche che, benché metodologicamente scorretti, potrebbero essere fatti. Il secondo elemento rende altrettanto sbagliati i confronti tra SSD basati sulla *peer review*, confronti che però sarebbe utile poter fare.

Il GEV 11 ritiene perciò opportuno arrivare non all'abbandono o al superamento della *peer review* come strumento di valutazione *ex-post*, ma piuttosto:

- a. Ad un suo affinamento ottenuto attraverso una maggiore sofisticazione e differenziazione della scheda di valutazione, e l'uso di un corpo di valutatori più esperto e meglio selezionato in base all'esperienza (la VQR lascia a questo proposito una eredità importante, costituita dalla banca dati di *referee* prima scelti con cura e poi ulteriormente selezionati sulla base di un'esperienza che ne ha aumentato la competenza e la sensibilità, e da una migliore comprensione dei criteri in base ai quali scegliere nuovi revisori);
- b. Ad una sua integrazione con altri strumenti di valutazione, ottenibili per esempio con la costruzione di uno strumento bibliometrico ausiliario, specifico per le discipline umane e sociali italiane. Il rischio è naturalmente che tale strumento possa trasformarsi in una sorta di incoraggiamento e difesa dell'autarchia culturale da cui talvolta queste discipline sono afflitte. E' un rischio che è però possibile ridurre pesando opportunamente i risultati ottenuti con questo strumento rispetto a quelli ottenuti, per gli stessi autori, utilizzando le grande banche dati internazionali come ISI, in base all'innegabile constatazione che la ricerca italiana è, anche in questi campi, solo una parte della *république des lettres* internazionale, e che quindi in quanto parte deve essere pensata.

In alternativa si potrebbe pensare all'adozione per il futuro di un modello simile al *Research Excellence Framework* britannico, con *panel* più piccoli e specializzati (supponendo che la Psicologia continui ad affidarsi alla bibliometria, l'Area 11 dovrebbe in quest'ottica averne almeno cinque: Antropologia, Filosofia, Geografia, Pedagogia e Storia), composti da una trentina di studiosi ciascuno, posti in congedo per qualche mese, che insieme valutino omogeneamente campioni significativi di prodotti. E' però forte il pericolo che tale scelta non elimini i rischi di soggettività, e anzi ne rafforzi probabilità e dimensioni, visto che si tratterebbe di affidare a poche persone la valutazione di intere, e variegate, aree disciplinari.

Se insomma nessun sistema di valutazione è perfetto, come dimostra il fatto che ogni paese dopo averne sperimentato uno è spesso ritornato sui suoi passi per cambiarlo o migliorarlo,



si tratta di puntare costantemente al loro affinamento, e non di usare le sempre inevitabili imperfezioni per propugnare il ritorno ad un'epoca in cui la valutazione era in realtà molto più primitiva e permetteva l'esistenza di ampie e costose aree di stagnazione e emarginazione, nonché arbitrii spesso slegati da considerazioni scientifiche.

La VQR ha fatto poi emergere anche in Area 11 il problema della valutazione della professionalità. L'Università non è infatti solo ricerca, e didattica, o meglio unione tra ricerca e didattica, ma anche luogo dell'alta formazione professionale. La VQR ha giustamente tenuto conto della sola ricerca, domani una VQDidattica valuterà invece quest'ultima, certo in rapporto con la prima. Bisognerà però arrivare anche a una valutazione della professionalità e quindi della capacità di garantire formazione professionale ai massimi livelli: quella di un bravo psicologo, di un bravo chirurgo o di un grande avvocato, non si misura infatti solo con la sua capacità di produrre articoli scientifici o di tenere lezioni brillanti. Le Università andranno insomma valutate con uno spettro di strumenti diversi, e potranno così trovare anche grazie ad essi la loro vocazione differenziata.

Andrea Graziosi, Presidente

Approvato telematicamente dal GEV 11 il 20 giugno 2013