

# REPORT:

## Pratiche basate sui dati nella didattica universitaria: il caso dell'Università di Padova

*Juliana E. Raffaghelli, Valentina Grion, Marina De Rossi*

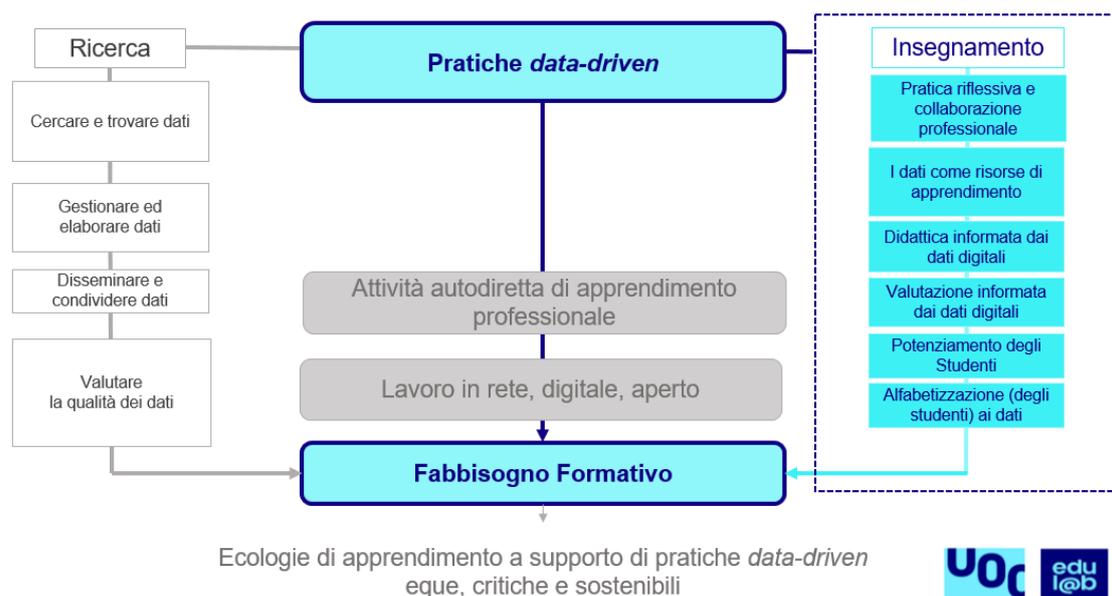
### 1. Introduzione

Il presente studio si basa su un'indagine condotta a livello interateneo, a partire dalla collaborazione nata nel seno di un progetto di ricerca nazionale finanziato dal Ministero dell'Innovazione e la Ricerca di Spagna (Professional Learning Ecologies for Digital Scholarship: Modernizing Higher Education by Supporting Professionalism: <http://edulab.uoc.edu/en/projects/led-projects/professional-learning-ecologies-for-digital-scholarship-modernizing-higher-education-by-supporting-professionalism/> ). Il lavoro coordinato ha avuto come motivazione la comune necessità di comprendere in profondità le pratiche e nodi critici relativi all'uso dei dati, come componente emergente della competenza digitale nel contesto della professione accademica nell'era digitale. Ad una fase di *desk research* mirante alla definizione di costrutti e l'orientamento delle domande di ricerca, è seguita la costruzione di un questionario. L'università di Padova ha partecipato sia alla validazione teorica, sia empirica. La fase successiva è stata infatti l'implementazione di un'indagine atta a fornire le informazioni necessarie per elaborare gli scenari e il quadro delle competenze per l'alfabetizzazione ai dati per la didattica universitaria.

La Figura 1 presenta i costrutti principali adottati per la costruzione dello strumento (questionario). Nella fattispecie, sono stati utilizzati tre costrutti principali: a) il profilo del docente universitario (genere, età, esperienza professionale in ricerca e didattica, settore scientifico-disciplinare) b) il quadro di competenze per le pratiche basate su dati nella didattica universitaria; c) le modalità di apprendimento ovvero *ecologie di apprendimento* tra formale e informale (Sangrá, Raffaghelli, & Guitert-Catasús, 2019).

**Figura 1**– Costrutti adottati nello studio (figura totale). In colore, i costrutti analizzati e comunicati in questo articolo.

## Professione docente nel l'era digitale: le pratiche basate sui dati digitali



Considerata la densità teorica dei componenti b) e c), la comunicazione scientifica dei risultati è stata organizzata in due fasi, di cui quest'articolo ne riporta la prima, ovvero, quella relativa all'analisi del quadro attuale di competenze per l'uso dei dati nella didattica universitaria, che supporta l'analisi del fabbisogno formativo. Come si evince dalla figura 1, i costrutti coprivano tutte le componenti della *digital scholarship* come presentata e discussa nello stato dell'arte di cui sopra; nonché l'identificazione delle ecologie di apprendimento.

La tabella 1 mostra i costrutti di ricerca e le domande di ricerca connesse.

**Tabella 1** – Costrutti adottati nello studio e Domande di ricerca

Aree di pratiche basate sull'uso di dati digitali nella didattica universitaria	N di Items	Domande di Ricerca
<p><b>EDMQ (Educational Data use for Management and Quality) - Uso di dati per la gestione e la qualità della didattica</b>                      Uso dei dati da diverse fonti per informare i processi istituzionali, la collaborazione e la costruzione di conoscenza accademica a supporto della didattica, inclusa la ricerca.</p>	<p><b>8 item</b>                      EDMQ1 Uso di dati da Report Nazionali per lo sviluppo istituzionale e la pianificazione                      EDMQ2 Uso di dati da Report Istituzionali per lo sviluppo istituzionale e la pianificazione                      EDMQ3 Uso di dati dalla valutazione del proprio corso per lo sviluppo istituzionale e la pianificazione                      EDMQ4 Uso di dati del proprio corso per la progettazione curricolare                      EDMQ5 Uso di analitiche di apprendimento per la progettazione didattica                      EDMQ6 Uso di analitiche di apprendimento per analizzare efficacia didattica propria                      EDMQ7 Uso di dati da social media inseriti nel mio corso per analizzare la propria efficacia didattica                      EDMQ8 Uso di dati da social media dove partecipano gli studenti in generale per la progettazione e l'efficacia didattica</p>	<p>Quali sono le pratiche basate sull'uso di dati digitali nella didattica più diffuse all'interno dell'Ateneo di Padova?</p> <p>Esistono differenze attraverso le varie tipologie di pratica, considerando particolarmente l'ambito disciplinare?</p>
<p><b>DER (Data as Educational)</b></p>	<p><b>4 item</b></p>	

Pratiche basate sui dati nella didattica universitaria

<p><b>Resources) - Uso dei dati come risorse di apprendimento.</b> Adozione di dati provenienti dalla propria ricerca o dati aperti pubblici per orientare e supportare attività di apprendimento.</p>	<p>DER1 Uso di dati di report internazionali/nazionali come risorsa didattica DER2 Riutilizzo di Dati Aperti (Open Data) provenienti dalla ricerca o dalla pubblica amministrazione come risorsa didattica DER3 Riutilizzo di Dati Aperti (Open Data) provenienti dalla propria ricerca come risorsa didattica DER4 Riflessione su aspetti etici relative all'integrazione dei dati aperti come risorse didattiche.</p>	
<p><b>DTL (Data supporting Teaching and Learning) - Didattica informata dai dati digitali.</b> Decisione docente, monitoraggio di processi di apprendimento e feedback con uso di dati digitali dai sistemi adottati per la didattica.</p>	<p><b>5 item</b> DTL1 Integrazione di modalità tradizionali di raccolta dati come parte delle attività e compiti degli studenti. DTL2 Integrazione di tecniche estrattive di dati (massivi) come parte delle attività e compiti degli studenti. DTL3 Insegnamento di tecniche per l'elaborazione/ rappresentazione/ visualizzazione di dati DTL4 Riflessione docente sugli aspetti di estrazione automatica dei dati per utilizzo didattico DTL5 Integrazione di dashboards (pannelli di controllo visuali) per supportare le abilità autoregolative degli studenti.</p>	
<p><b>DA (Data supporting Assessment) - Valutazione informata dai dati digitali.</b> Decisione docente e analisi dei percorsi formativi compiuti a partire da dati ottenuti/estratti tramite piattaforme digitali.</p>	<p><b>10 item</b> DA1 Uso di dati ottenuti da verifiche progressive per monitoraggio dell'apprendimento DA2 Uso di dati ottenuti da verifiche progressive per monitoraggio didattico DA3 Uso dei dati ottenuti da verifiche per dare feedback sommativi agli studenti DA4 Uso dei dati ottenuti dalla valutazione di un corso per dare feedback formativi agli studenti DA5 Riflessione insieme agli student sui processi di raccolta dei dati provenienti dalla didattica. DA6 Uso dei dati dalle piattaforme di apprendimento per monitorare/valutare la didattica. DA7 Uso di "dashboards" presenti nelle piattaforme di apprendimento per informare la mia didattica DA8 Uso di sistemi digitali automatizzati per analizzare e valutare i risultati degli studenti (e.g. online quiz) DA9 Uso di dati dalle piattaforme di apprendimento per riflettere insieme agli studenti sulla qualità della didattica. DA10 Uso di sistemi digitali automatizzati per valutare in modo complessivo la qualità della didattica (e.g., online surveys).</p>	
<p><b>SED (Students' Empowerment through Data) - Potenziamento degli Studenti.</b> Coinvolgimento degli studenti nella comprensione della propria partecipazione e dei propri processi di apprendimento a partire dagli ecosistemi</p>	<p><b>3 item</b> SED1 Condivisione di report istituzionali per supportare la riflessione degli student sul progresso dei propri apprendimenti nel contesto istituzionale o sociale. SED2 Condivisione di dati provenienti dalla piattaforma di apprendimento (analitiche di apprendimento) per supportare la riflessione degli student sul progresso dei propri apprendimenti nel contesto istituzionale o sociale. SED3 Discussione su analitiche estratte da social media (integrati nella didattica) con gli student, ove</p>	

di dati generati dagli ambienti di didattica a distanza.	consentire loro di riflettere sul progresso dei propri apprendimenti nel contest istituzionale o sociale.	
<b>SDL (Students' Data Literacy) - Alfabetizzazione (degli studenti) ai dati.</b> Azioni didattiche mirate allo sviluppo di conoscenza e skills tecniche, grafiche, multimediali, critiche ed estetiche sull'uso dei dati nei contesti di vita e professionali.	<b>5 item</b> SDL1 Promozione dell'abilità degli studenti per valutare criticamente la affidabilità e credibilità dei dati incorporati in una fonte informativa. SDL2 Promozione di skills tra gli studenti per integrare dati in modo significativo entro report narrativi o presentazioni SDL3 Promozione della discussione e riflessione tra gli studenti sui vincoli sociali relativi all'uso dei dati e le applicazioni digitali nell'IA come processo socio-tecnico. SDL4 Promozione dell'analisi etico relativo raccolta e utilizzo dei propri dati come parte del processo di apprendimento. SDL5 Discussione congiunta con gli studenti sull'impatto dell'uso dei dati nella società in generale.	

Lo strumento, nella fattispecie un questionario di 21 domande, è stato sottoposto a due tipi di validazione.

La prima è stato uno studio Delphi (Raffaghelli, 2019a) a partire dal quale le dimensioni del questionario sono state consolidate. I feed-back dati da 8 esperti partecipanti nelle due fasi dello studio Delphi sono stati analizzati sia a livello qualitativo, sia attraverso procedure quantitative. Dal punto di vista qualitativo, le osservazioni degli esperti sono risultate utili a consolidare le dimensioni di analisi dell'alfabetizzazione ai dati, nonché delle ecologie di apprendimento, teoricamente definite. Dal punto di vista quantitativo, gli indici Kappa di Fleiss e W di Kendall, sono stati utilizzati per misurare l'associazione e accordo tra i giudizi dati dagli esperti. I valori ottenuti sono da ritenersi rilevanti ( $K=.84$ ,  $W=.86$ ), indicando un generale accordo tra esperti sugli item da tenere e/o modificare.

In secondo luogo, è stata portata avanti un'analisi empirica di affidabilità attraverso il test de Cronbach. L'indice  $\alpha$  de Cronbach è stato calcolato per tutte le variabili sotto analisi, ottenendosi un  $\alpha = .94$  (molto alto) per ciascuna delle variabili di risposta (pratiche di dati), ottenendosi i seguenti valori: EDMQ  $\alpha = .8$ ; DER  $\alpha = .803$ ; DTL  $\alpha = .0.793$ ; DA  $\alpha = .0.876$ ; SED  $\alpha = .0.764$ ; SDL  $\alpha = .0.853$ . I valori, tutti vicini o sopra lo 0.8 sono da ritenersi buoni, essendo la soglia di accettabilità i valori al di sotto di 0.70. Tali risultati indicano che le scale riescono a misurare con accuratezza accettabile a buona le variabili d'interesse (Cortina, 1993).

## 2. Raccolta Dati e Partecipanti

La raccolta dati è stata portata avanti attraverso questionario online, validato in fase previa a partire da uno studio Delphi (citazione). Il questionario è stato inviato tramite mailing list dell'università a tutti i docenti (professori ordinari e associati), ricercatori e assegnisti dell'università di Padova (4464 partecipanti). Una disamina dei dati forniti da USTAT 2017/2018 (<http://ustat.miur.it/dati/didattica/italia/atenei-statali/padova>) che, con 57.927 studenti e 4464 docenti, l'università di Padova si colloca tra le più grandi università

Pratiche basate sui dati nella didattica universitaria

pubbliche italiane del Nord (media studenti universitari italiani in atenei pubblici: 21.994; media dei docenti universitari negli atenei pubblici italiani: 1237). Il numero di docenti coinvolti nello studio rappresenta quindi il 4% del totale dei docenti universitari italiani. In totale, il tasso di risposta al questionario online è stato del 8%.

Per quanto riguarda il margine d'errore del campione, per un intervallo di confidenza del 95% (tradizionalmente utilizzato nelle scienze sociali) e un'eterogeneità massima dello  $p=q=.05$ , si ottiene quindi un valore di  $\pm 5.09\%$  per tutta popolazione docente degli atenei italiani, e dello  $\pm 4.88\%$  per l'ateneo patavino.

Benché si tratti di un campione incidentale e volontario, non-probabilistico, il numero di docenti partecipanti rappresentano consistentemente il caso dell'ateneo patavino e delle grandi università pubbliche italiane. Tuttavia, generalizzeremo i risultati ottenuti al solo caso dell'Università di Padova in quanto ecosistema di apprendimento professionale.

### 3. Procedure di analisi dei dati

I dati sono stati trattati in due fasi. Nella prima fase, sono state elaborate statistiche descrittive delle variabili sotto studio, ovvero caratteristiche del campione in termini di età, genere e ambito scientifico-disciplinare, e pratiche basate sui dati nella didattica universitaria (cfr. Tabella 1, EDMQ, DER, DTL, DA, SED, EDL). Le caratteristiche del campione compongono un primo gruppo di tre variabili categoriali in tutto, che sono modulatori delle pratiche basate sui dati nella didattica in quanto secondo gruppo di variabili ordinali (risposte secondo scala Likert). Ove consentire le procedure di analisi, il secondo gruppo è stato trasformato in variabili numeriche discontinue (scala 1-5). Inoltre, sono state analizzate le caratteristiche delle distribuzioni di ciascuna delle variabili sotto analisi, utilizzando frequenze, percentuali e test Chi-Square di Pearson per l'analisi delle distribuzioni di variabili categoriali; e misure di tendenza centrale (media, mediana) e dispersione (deviazione standard, quartili), nonché curtosi e asimmetria per le variabili numeriche.

Successivamente, si è applicata una tecnica di riduzione dimensionale per analizzare l'eventuale esistenza di dimensioni latenti inerenti ai vari item dentro ogni area di pratica basata sui dati nella didattica universitaria (EDMQ=8 items, DER=4 items, DTL=5, DA=10, SED=3, EDL=5). La tecnica scelta è stata la *Principal Components Analysis* (PCA). La tecnica si basa nell'identificazione di componenti principali a partire da un processo di calcolo sulle osservazioni (punti nello spazio vettoriale di relazione tra variabili) che quindi proietta ogni punto dati solo sulle componenti principali (vettori) per ottenere dati di dimensioni inferiori preservando il più possibile la variazione dei dati. In pratica, ciò che si ottiene è un numero minore di variabili associate sotto una componente in una matrice di covarianza. Le componenti possono quindi essere utilizzate per successive interpretazioni e inferenza statistica. La PCA è la più semplice delle vere analisi multivariate basate su autovettori ed è strettamente correlata all'analisi fattoriale, sebbene quest'ultima può incorporare ipotesi più specifiche del dominio sulla struttura sottostante. Considerando la complessità di domande ritenute teoricamente rilevanti nel nostro studio iniziale tramite questionario, la PCA potrebbe aiutare a svelare strutture di relazioni tra variabili e quindi a caratterizzare le pratiche basate sui dati in modo più

Pratiche basate sui dati nella didattica universitaria

sintetico. Data la natura esplorativa del nostro studio, basato in domande di ricerca che puntano a comprendere la portata del fenomeno, la PCA è stata ritenuta la tipologia di analisi statistica più opportuna.

## 4. Risultati

### 4.1. Caratteristiche del Campione

Il campione mostra una distribuzione non omogenea in relazione alle appartenenze disciplinari, l'età e il genere. L'osservazione della Tabella 2 consente di osservare che vi è una chiara prevalenza di casi dell'ambito delle scienze naturali seguito dalle scienze sociali (che attestano quasi il 50% del campione), anche se poi la distribuzione si stabilizza in relazione all'ambito Umanistico insieme al ambito delle Tecnologie (incluse le ingegneria), con un quarto dei casi appartenenti a questi settori. Rispetto alle fasce di età, prevalgono i casi sopra i 45 anni (45-55, 29% e > 55, 35%). Meno della metà si colloca al di sotto di quelle due fasce di età (<25, 25-34 e 35-44 rappresentano il 25% del campione). Infine, per quanto riguarda il genere, oltre 2 casi non dichiarati e 40 risposte assenti, si osserva una netta prevalenza di maschi (54%) rispetto delle femmine (35%). I valori del test chi-quadrato confermano la prevalenza significativa (p-value < .01) di categorie nella distribuzione. In tutto, si può sostenere che la distribuzione non si discosta dall'universo nazionale italiano, motivo per il quale, sebbene non ci sono le basi probabilistiche del campione per una generalizzazione di risultati, gli stessi sono da ritenersi un caso rilevante.

**Tabella 2** – Partecipanti Totali per Ambito Disciplinare, Età e Genere

	NA	Nat.Sci	Soc.Sci	Hum	Formal. Sci	Biomed	Tech	Totale
Freq	42	98	89	47	27	26	41	370
%	11,35%	26,49%	24,05%	12,70%	7,30%	7,03%	11,08%	100,00%
X2 test GOF								
$\chi^2 = 158.43$ , df = 7, p-value < 2.2e-16								
	NA	< 25	25-34	35-44	45-54	> 55	Totale	
Freq	39	1	12	81	108	129	370	
%	10,54%	0,27%	3,24%	21,89%	29,19%	34,86%	100,00%	
X2 test GOF								
$\chi^2 = 222.41$ , df = 5, p-value < 2.2e-16								
	NA	Femmine	Maschi	No Dichiarato	Totale			
Freq	40	130	198	2	370			
%	10,81%	35,14%	53,51%	0,54%	100,00%			
X2 test GOF								
$\chi^2 = 253.87$ , df = 3, p-value < 2.2e-16								

La Tabella 3 presenta una statistica descrittiva bivariata, dove è di nuovo possibile cogliere alcune costanti nella distribuzione tra ambito disciplinare, età e genere. In particolare, si noti la ampia prevalenza maschile nelle aree STEM e nelle fasce di età sopra i 45 anni (tecnologie, genere maschile 7% , genere femminile 2%; scienze naturali,

Pratiche basate sui dati nella didattica universitaria

maschile 11%, femminile 8%; scienze formali, maschile 4%, femminile 1%). Sbilanciamento che si osserva anche nelle scienze sociali (maschi 9%, femmine 6%) e ambito umanistico (maschile 7%, femminile 4%). Solo nelle scienze biomediche si trova una situazione inversa, con l'1% di maschi e il 3% femmine). È anche in quest'ultimo settore che si attesta una maggiore presenza di giovani (sotto i 45 anni) con un 2% di femmine e 0.3% di maschi. Da considerare il campione autoselezionato, con prevalenza di risposte da settori socio-sanitari altamente femminilizzati (come quello infermieristico e quello della nutrizione). I settori disciplinari in cui vi è maggiore presenza giovanile (sotto i 35 anni) è quello delle scienze sociali per le femmine (0.8%) e quello delle scienze naturali per il genere maschile (1.1%).

**Tabella 3** – Analisi bivariato (Età e Ambito Disciplinare per Genere)\*

\*NA=8.91, Maschile + Femminele = 91.09%

<b>FEMMINILE</b>														
<b>Ambito Disc.</b>	Età													
	< 25		25-34		35-44		45-54		> 55		Totale			
	Fr	%	Fr	%	Fr	%	Fr	%	Fr	%	Fr	%		
NA	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Nat.Sci	0	0	0	0	1	0,27	8	2,16	13	3,51	15	4,05	37	10,00
Soc.Sci	0	0	0	0	3	0,81	12	3,24	16	4,32	8	2,16	39	10,54
Hum Formal.	0	0	0	0	1	0,27	3	0,81	9	2,43	7	1,89	20	5,41
Sci	0	0	0	0	0	0,00	3	0,81	3	0,81	1	0,27	7	1,89
Biomed	0	0	0	0	0	0,00	9	2,43	3	0,81	8	2,16	20	5,41
Tech	0	0	0	0	0	0,00	1	0,27	5	1,35	1	0,27	7	1,89
Totale Genere Femminile 35.14%														
<b>MASCHILE</b>														
<b>Ambito Disc.</b>	Età													
	< 25		25-34		35-44		45-54		> 55		Totale			
	Fr	%	Fr	%	Fr	%	Fr	%	Fr	%	Fr	%		
NA	0	0	0	0	0	0	0	0	0	2	0	0	0	0
Nat.Sci	0	0	0	0	4	1,08	14	3,78	15	4,05	27	7,30	60	16,22
Soc.Sci	0	0	0	0	2	0,54	11	2,97	17	4,59	20	5,41	50	13,51
Hum Formal.	0	0	0	0	0	0,00	7	1,89	8	2,16	22	5,95	37	10,00
Sci	0	0	0	0	0	0,00	6	1,62	6	1,62	8	2,16	20	5,41
Biomed	0	0	0	0	0	0,00	1	0,27	2	0,54	3	0,81	6	1,62
Tech	0	0	0	0	1	0,27	6	1,62	11	2,97	16	4,32	34	9,19
Totale Genere Maschile 55.95%														

La tabella 4 introduce la distribuzione dei punteggi relative alle scale di pratiche basate sui dati nella didattica universitaria, ovvero le 6 scale con tipologia di risposta Likert (range 1-5, esprimendo totale disaccordo a totale accordo). In questo caso la curtosi e asimmetria indicano distribuzioni normali, in alcuni casi con una tendenza negativa legata a valori di risposta bassi (nella nostra scala, bassa frequenza di una pratica basata sui dati).

Da notare che la scala EDMQ (Uso dei dati per la gestione e la qualità della didattica) è quella che attesta i valori medi più alti rispetto alle altre 5 scale. In particolare, sembra che tra i rispondenti ci sia una pratica diffusa del uso dei dati delle piattaforme per riprogettare la didattica (3.73, valore vicino all'accordo); e del uso dei dati del proprio corso per la progettazione curricolare (3.58). Molto bassi i punteggi relativo all'uso dei

Pratiche basate sui dati nella didattica universitaria

dati da social media inseriti nel proprio corso (EDMQ7) e dai social media in generale (EDMQ8) come fonti informative usate per l'analisi della qualità della didattica e la riprogettazione.

L'altro unico caso in cui il valore raggiunge il punteggio di 3 (né in disaccordo né d'accordo) che ad ogni modo indica incertezza, è quello della scala DA (Valutazione informata dai dati digitali). Infatti, i docenti indicano una eventuale frequenza di uso dei dati ottenuti da verifiche progressive per il monitoraggio didattico (DA2, 3.02) e dei dati ottenuti da verifiche per dare feedback sommativi agli studenti (DA3, 2.90). Da considerare anche il valore di SDL1 (promozione dell'abilità degli studenti di valutare criticamente l'affidabilità e credibilità dei dati incorporati in una fonte informativa) presente in qualche modo come pratica a partire dal valore  $2.90 \pm 1.73$ ; nonché la promozione di skills tra gli studenti per integrare dati in modo significativo entro report narrativi o presentazioni ( $2.90 \pm 1.74$ ).

**Tabella 4-** Distribuzione delle risposte quantitative (Scala Likert 1-5)

Var	Media	Dev.ST	Min	Q1	Mediana	Q3	Max	MAD	IQR
EDMQ1	1.71	1.43	0.00	1.00	1.00	2.00	5.00	1.48	1.00
EDMQ2	2.38	1.58	0.00	1.00	2.00	4.00	5.00	1.48	3.00
EDMQ3	3.28	1.63	0.00	2.00	4.00	5.00	5.00	1.48	3.00
EDMQ4	<b>3.58</b>	1.49	0.00	3.00	4.00	5.00	5.00	1.48	2.00
EDMQ5	<b>3.73</b>	1.46	0.00	3.00	4.00	5.00	5.00	1.48	2.00
EDMQ6	1.84	1.68	0.00	1.00	1.00	3.00	5.00	1.48	2.00
EDMQ7	<b>1.35</b>	1.43	0.00	0.00	1.00	2.00	5.00	1.48	2.00
EDMQ8	<b>0.99</b>	1.09	0.00	0.00	1.00	1.00	5.00	1.48	1.00
DER1	2.08	1.68	0.00	1.00	2.00	4.00	5.00	1.48	3.00
DER2	2.46	1.67	0.00	1.00	2.00	4.00	5.00	1.48	3.00
DER3	2.44	1.66	0.00	1.00	2.00	4.00	5.00	1.48	3.00
DER4	2.05	1.65	0.00	1.00	2.00	3.00	5.00	1.48	2.00
DTL1	2.12	1.75	0.00	1.00	2.00	4.00	5.00	1.48	3.00
DTL2	1.11	1.22	0.00	0.00	1.00	1.00	5.00	1.48	1.00
DTL3	1.88	1.69	0.00	1.00	1.00	3.00	5.00	1.48	2.00
DTL4	1.65	1.56	0.00	0.00	1.00	2.00	5.00	1.48	2.00
DTL5	1.75	1.70	0.00	0.00	1.00	3.00	5.00	1.48	3.00
DA1	2.89	1.65	0.00	1.00	3.00	4.00	5.00	1.48	3.00
DA2	<b>3.02</b>	1.66	0.00	2.00	3.00	5.00	5.00	1.48	3.00
DA3	<b>2.90</b>	1.67	0.00	1.00	3.00	4.00	5.00	1.48	3.00
DA4	2.67	1.66	0.00	1.00	3.00	4.00	5.00	1.48	3.00
DA5	2.32	1.62	0.00	1.00	2.00	4.00	5.00	1.48	3.00
DA6	1.71	1.54	0.00	1.00	1.00	3.00	5.00	1.48	2.00
DA7	<b>1.18</b>	1.26	0.00	0.00	1.00	1.00	5.00	1.48	1.00
DA8	1.69	1.61	0.00	1.00	1.00	3.00	5.00	1.48	2.00
DA9	<b>1.16</b>	1.15	0.00	0.00	1.00	1.00	5.00	1.48	1.00
DA10	1.56	1.62	0.00	0.00	1.00	2.00	5.00	1.48	2.00
SED1	1.69	1.52	0.00	1.00	1.00	3.00	5.00	1.48	2.00
SED2	<b>2.09</b>	1.58	0.00	1.00	2.00	3.00	5.00	1.48	2.00
SED3	<b>1.07</b>	1.21	0.00	0.00	1.00	1.00	5.00	1.48	1.00
SDL1	<b>2.90</b>	1.73	0.00	1.00	3.00	4.00	5.00	2.97	3.00
SDL2	2.48	1.73	0.00	1.00	2.00	4.00	5.00	2.97	3.00
SDL3	<b>1.38</b>	1.39	0.00	0.00	1.00	2.00	5.00	1.48	2.00
SDL4	1.52	1.43	0.00	1.00	1.00	2.00	5.00	1.48	1.00
SDL5	1.87	1.54	0.00	1.00	1.00	3.00	5.00	1.48	2.00

Alcune scale registrano maggiore incertezza, con valori che tuttavia lasciano intuire una eventuale attivazione, anche se discontinua, verso la pratica basata su dati. È il caso dell'uso di dati come risorsa educativa (scala DER, in particolare items 1,2,3 con mediana 3); l'integrazione di modalità tradizionali di raccolta dati come parte delle attività e compiti degli studenti, ovvero compiti autentici degli studenti basati sui dati (DTL1, mediana 3); e le pratiche di uso dei dati nel monitoraggio e la valutazione dell'apprendimento (DA 1-5 con mediana 3, item relativi a pratiche relative all'uso di dati ottenuti attraverso verifiche progressive come base per il monitoraggio e il feedback sommativo agli studenti, nonché per supportare la riflessione relativa alla struttura di valutazione dell'apprendimento all'interno del corso).

Molto scarso l'uso dei dati sia per il potenziamento degli studenti (Scala SED, con valori medi vicini o sotto il 2) che per la promozione di abilità tecniche e critico-etiche come parte di processi di alfabetizzazione in dati. Infatti, si osservava bassa o nessuna frequenza di uso di tecniche avanzate di data mining come parte delle attività per gli studenti (DTL2, 1.11); bassa frequenza di attività di promozione dell'analisi etico relativo alla raccolta e utilizzo dei propri dati all'interno del processo di apprendimento (SDL4, 1.52) e nella società in generale (SDL5, 1.87). In generale, tutte le pratiche basate sui dati che necessitano di operazioni *data-driven* con uso di risorse e concetti metodologici più recenti, hanno mostrato valori più bassi. Si tratta di tutti gli ultimi 2 a 4 item delle 6 scale, come per esempio DTL5, 1.56, integrazione di dashboards per supportare abilità autoregolate degli studenti; oppure D10, uso di sistemi digitali automatizzati per valutare in modo complessivo la qualità della didattica. Da notare che tali valori sono profondamente legati alla disponibilità di strutture tecnologiche, nonché di un accurato supporto tecnico e pedagogico per poter sperimentare *in vivo* con gli studenti o a livello di analisi docente, tali operazioni.

**Tabella 5** - Assimetria e Curtosi

I valori di asimmetria e curtosi compresi tra -2 e +2 sono considerati accettabili al fine di dimostrare una normale distribuzione univariata. I valori tra -.5 e .5 indicano una distribuzione approssimativamente simmetrica; i valori tra -1 e -.5, così come tra .5 e 1, indicano una distribuzione moderatamente asimmetrica. In valori di asimmetria negativi indicano una maggiore presenza di risposte a valori bassi (1-2, frequenza bassa o nulla delle pratiche basate sui dati) mentre i valori di asimmetria positiva indicano una maggiore presenza di risposte a valori alti (4-5, frequenza alta o molto alta di pratiche basate sui dati)

Var	Skewness	SE.Skew	Kurtosis	N.Valid	Pct.Valid
EDMQ1	0.89	0.15	-0.11	255.00	68.92
EDMQ2	0.29	0.15	-1.05	255.00	68.92
EDMQ3	-0.54	0.15	-0.97	255.00	68.92
EDMQ4	-0.76	0.15	-0.58	255.00	68.92
EDMQ5	-0.99	0.15	-0.08	255.00	68.92
EDMQ6	0.62	0.15	-0.98	254.00	68.65
EDMQ7	1.25	0.15	0.69	253.00	68.38
EDMQ8	1.59	0.15	2.82	254.00	68.65
DER1	0.48	0.15	-1.10	255.00	68.92
DER2	0.07	0.15	-1.25	255.00	68.92
DER3	0.10	0.15	-1.27	255.00	68.92
DER4	0.44	0.15	-1.08	255.00	68.92
DTL1	0.40	0.15	-1.25	255.00	68.92
DTL2	1.53	0.15	2.06	255.00	68.92

Pratiche basate sui dati nella didattica universitaria

DTL3	0.63	0.15	-0.88	255.00	68.92
DTL4	0.87	0.15	-0.34	255.00	68.92
DTL5	0.80	0.15	-0.69	255.00	68.92
DA1	-0.20	0.15	-1.28	255.00	68.92
DA2	-0.35	0.15	-1.15	255.00	68.92
DA3	-0.24	0.15	-1.28	255.00	68.92
DA4	-0.07	0.15	-1.27	255.00	68.92
DA5	0.24	0.15	-1.10	255.00	68.92
DA6	0.83	0.15	-0.39	255.00	68.92
DA7	1.38	0.15	1.39	255.00	68.92
DA8	0.87	0.15	-0.46	255.00	68.92
DA9	1.49	0.15	2.32	255.00	68.92
DA10	0.98	0.15	-0.32	255.00	68.92
SED1	0.80	0.15	-0.47	255.00	68.92
SED2	0.43	0.15	-0.88	255.00	68.92
SED3	1.61	0.15	2.42	255.00	68.92
SDL1	-0.24	0.15	-1.37	255.00	68.92
SDL2	0.06	0.15	-1.34	255.00	68.92
SDL3	1.19	0.15	0.54	255.00	68.92
SDL4	1.01	0.15	0.15	255.00	68.92
SDL5	0.55	0.15	-0.83	255.00	68.92

## 4.2. Analisi delle Componenti Principali (PCA)

In seguito, presentiamo l'analisi delle scale con estrazione di componenti principali per la riduzione della dimensionalità e identificazione di variabili latenti che attestano la presenza di fenomeni complessi relativi all'uso dei dati digitali come parte della didattica universitaria nel caso analizzato.

La tecnica di analisi di PCA richiede step successivi anche per l'eventuale correzione del modello di variabili latenti emergenti. Data la mole di dati emergenti, gli stessi sono presentati negli allegati II (matrici di correlazione, che consentono di intuire le relazioni tra items della scala) e III (identificazione delle componenti e grafici *screeplot* per rappresentare tale identificazione). Questo III allegato infatti si compone da: a) le tabelle di sintesi di PCA, dove si può comprendere la contribuzione delle variabili in ciascuna delle componenti e l'importanza relativa delle componenti nell'esplicare la varianza del *subset* di dati analizzati (ovvero la scala di uso dei dati nella didattica); b) gli *eigenvalues* o autovalori che identificano la quantità di componenti principali; c) i *loadings* o *pesi alle variabili*, che identifica la contribuzione di ciascuno degli item della scala ai componenti identificati; d) la rotazione ortogonale *promax* che identifica relazioni tra i componenti) ed e) la rappresentazione grafica di *Screeplot* che rende evidente la quantità di componenti trovati. In questa sede commentiamo e riportiamo i valori all'interno del testo per semplificare la lettura. Infine, vengono forniti grafici che rappresentano in modo più succinto la contribuzione degli item e numero di componenti o dimensioni (Figura 1); e la rappresentazione "*biplot*" che presenta gli item della scala come vettori in uno spazio bidimensionale tra i due principali componenti identificati. La lunghezza dei vettori indica il peso del item rispetto al componente, e la direzione, la relazione con il primo o secondo componente. In questo caso, abbiamo aggiunto la rappresentazione dei gruppi "ambito disciplinare" che può indicare eventuali differenze nella relazione tra componenti. Ogni componente va interpretato infatti a partire dalle variabili (o item) che lo costituiscono.

### 4.2.1 EDMQ: Uso di dati per la gestione e la qualità della didattica

Pratiche basate sui dati nella didattica universitaria

Di un totale di 8 items/variabili, per questa scala a partire dagli autovalori (eigenvalue) sono stati identificati tre componenti principali (Componente principal o CP1.1, 3.41; CP1.2, 1.54; CP1.3, 1.04). Lo *screeplot* infatti mostra di modo netto la presenza di queste tre componenti. La complessità media per item analizzata a partire da rotazione varimax e promax da come risultato 1.2, indicante la buona consistenza interna della componente (che rappresenta un fattore latente). La prima componente (CP1.1) spiega una percentuale di varianza del 42.74% mentre la seconda (CP1.2) del 19.34% e la terza (CP1.3) del 13.10%, arrivando a spiegare in totale il 75.2% della varianza. Da considerare che esiste un 25% non coperto, quindi la struttura delle componenti è accettabile, sebbene potrebbe migliorare in robustezza. Contribuiscono alla CP1.1. in particolare gli items EDMQ3, 4, 2, 1 (in questo ordine). Tale componente rappresenta maggiormente le pratiche basate sui dati più tradizionalmente associate alla valutazione della qualità, in particolare la consulta di report nazionali e internazionali, nonché istituzionali per orientare la progettazione curricolare e didattica. Sono aspetti che risultano evidenti per i docenti e ai quali sono probabilmente abituati attraverso le stesse strutture di governo istituzionale, come i piani strategici e le operazioni di autovalutazione della qualità istituzionale e didattica a partire da indicatori ANVUR. La seconda componente, che spiega una percentuale di variabilità molto meno rilevante (attorno al 20%) ha un peso relativamente importante dell'item EDMQ8 (uso dei dati da social media dove partecipano i miei studenti per la progettazione didattica). Da notare inoltre il peso negativo di EDMQ3 e 4 (ovvero che indica il comportamento opposto all'item sotto analisi all'interno della componente, e in questo caso ci si riferisce ai dati emersi dalla valutazione istituzionale del proprio corso e delle verifiche e valutazione degli apprendimenti all'interno del corso). Questa seconda componente sarebbe rappresentata da pratiche avanzate, che utilizzano più mezzi informali (come social media) per orientare le proprie attività di progettazione curricolare e di analisi della qualità della didattica. Rispetto alla correlazione tra componenti, si osserva giustamente una bassa correlazione tra le due prime, che spiegano comportamenti molto diversi (uso di dati istituzionalizzati rispetto ad uso di dati estratti da piattaforme digitali ad uso informale), e una maggiore correlazione tra prima e terza componente (la terza mostra valori molto deboli. La rappresentazione grafica *biplot* consente di vedere giustamente il comportamento opposto tra le due componenti. Aggiunge altresì una visione tra i vari ambiti disciplinari dove non si osservano differenze o raggruppamenti particolare (vi è una chiara dispersione dei punti/osservazioni nello spazio dimensionale). Si nota ad ogni modo che il gruppo delle Scienze Formali (che include matematica, statistica e computer science, mostra una tendenza positiva (e non opposta) per le due componenti. In effetti, questo gruppo potrebbe mostrare sistematicamente un uso avanzato di attività estrattive da piattaforme social media e open data, così come adderenza alle pratiche più tradizionali. Il gruppo Biomed è quello che mostra maggiore variabilità nelle risposte (elissi più ampia).

#### 4.2.2. DER: Uso dei dati come Risorse per l'apprendimento

In questo caso è stato individuato una sola componente principale, di un totale di 4 items/variabili: CP2.1, 2.52, confermato dallo *screeplot*. Si è considerata la seconda componente per la (CP2.2, 0.68) con lo scopo delle successive analisi rappresentazionali.

Pratiche basate sui dati nella didattica universitaria

La complessità media per item analizzata a partire da rotazione varimax e promax da come risultato 1, indicante ottima consistenza interna della componente (che rappresenta un solo fattore latente). La prima componente (CP2.1) spiega infatti una percentuale di varianza del 63.21% mentre la seconda (CP2.2) del 17.07%, in totale circa il 80% della varianza totale del subset di dati, da ritenersi buona. Contribuiscono alla CP2.1. in particolare gli items DER2 e DER3 (in questo ordine), che richiedono una progressiva comprensione dei dati aperti (Open Data) come risorsa educativa. Il DER2 indica le pratiche di riutilizzo di Dati Aperti privienienti dalla ricerca o dalla pubblica amministrazione come risorsa educativa, e DER3 il riutilizzo dei dati aperti generati dalla propria ricerca. La seconda componente spiega una percentuale di variabilità poco rilevante (17%). Tuttavia, è da considerare il peso dell'item DER4, relativo alla riflessione su aspetti etici connessi all'integrazione dei dati aperti come risorse didattiche. Di nuovo, sembra che in questa seconda componente si possano trovare gruppi di avanguardia che stanno già spigendo nella direzione di un uso critico e sapiente della ricchezza dei dati esistenti, in particolare Open Data. Andando a vedere la correlazione tra componenti, con uno 0.51% si attesta una accettabile correlazione tra la prima e seconda componente, che potrebbe indicare una buona predisposizione per passare dall'utilizzo o intenzionalità di utilizzo dei dati aperti alla riflessione etica e critica. La rappresentazione grafica *biplot* mostra una valenza molto negativa della componente DER4 per entrambi i componenti, e una moderata valenza positiva per la componente seconda di DER1 e DER2. Ciò conferma la scarsa frequenza di attivazione in relazione alla riflessione etico-critica sull'uso dei dati come risorsa educativa, benché vi sia una nascente tendenza all'utilizzo. Infine, andando a guardare le discipline, torniamo ad osservare un comportamento distaccato da parte del gruppo Formal Sciences che si colloca positivamente rispetto alle due componenti identificati, ovvero, non solo utilizza dati aperti e percorsi di data mining, ma si sta inoltrando nella considerazione di aspetti etici.

#### 4.2.3. DTL: Didattica informata dai dati digitali

La scala DTL centrava l'attenzione dei docenti universitari sui dati come alleati nel processo didattico, in particolare, la comprensione di dati prodotti dagli studenti per la presa di decisioni relative alla docenza. Gli autovalori (*eigenvalues*) hanno consentito di individuare una sola componente principale (CP3.1, 2.78) da un totale di 5 variabili. Una seconda componente non riesce a cogliere appieno la variabilità delle osservazioni, con un *eigenvalue* di 0.70. La complessità media per item analizzata a partire da rotazione varimax e promax da come risultato 1, indicante ottima consistenza interna della componente (che rappresenta un solo fattore latente). La percentuale di varianza spiegata dalle due prime componenti è di 55.73% per la prima componente (CP3.1) e del 14.10% per la seconda, in totale circa il 70% della varianza totale del subset di dati, da ritenersi accettabile. Il CP3.1 rispecchia l'impatto di tutti gli item, nel seguente ordine: DTL3,2,5,4. Tale componente può essere indicativa di una pratica consolidata di insegnamento di tecniche per l'elaborazione/rappresentazione/visualizzazione dei dati (DTL3) integrata all'implementazione sperimentale di tecniche estrattive come parte delle attività compito degli studenti. Eventualmente, i docenti potrebbero integrare dashboard di analitiche di apprendimento per supportare le abilità autoregolative degli studenti

Pratiche basate sui dati nella didattica universitaria

(DTL5) e riflettere insieme agli studenti (in casi eccezionali) sulle conseguenze dell'estrazione di dati degli studenti. La seconda componente può essere praticamente disdegnata (14% di variabilità). La correlazione tra componenti è interessante per quanto riguarda CP3.1. e CP3.4 e 5. Questi ultimi componenti non raggiungono la soglia per essere ritenuti componenti principali. Tuttavia, indicano come il CP3.1. (pratiche avanzate) si mette in relazione con le variabili con maggiore peso sui suddetti componenti (DTL2, DTL3 e DTL5) che insistono sull'integrazione di tecniche estrattive come attività compito; l'insegnamento di tecniche di visualizzazione e l'integrazione di dashboard. Si tratta sempre di pratiche avanzate nelle quali si potrebbe ipotizzare lo sforzo docente nell'applicare elementi di didattica in linea con il cambiamento di paradigmi epistemologico-scientifici, ovvero, *data-driven*. La rappresentazione grafica *biplot* mostra una valenza negativa per DTL4 e DTL5 nella relazione tra la prima e seconda componente; e una valenza piuttosto positiva per il secondo componente di DTL1. Tali valori indicano una scarsa influenza di attività riflessivo-partecipate sul componente, con una prevalenza su attività tradizionali di raccolta dati in quanto compito autentico degli studenti. Gli ambiti disciplinari non mostrano particolari raggruppamenti. Tuttavia, l'ambito Biomedico presenta maggiore dispersività dei dati, seguito dal gruppo delle Scienze Logico-formali. Entrambo i casi sembrano presentare una quantità di data points ben orientati verso metodologie *data-driven* condivise con gli studenti non solo a livello tecnico ma anche riflessivo.

#### 4.2.4. DA: Valutazione informata dai dati digitali

La scala in considerazione in questo paragrafo era la più lunga e complessa (con 10 item) tenendo conto il diffuso uso di dati analogici e digitali per l'assessment (monitoraggio, feedback e valutazione) ed il lungo percorso di ricerca e studio del problema delle metriche nella valutazione. Sono emerse in questo caso due componenti principali (CP4.1, *eigenvalue* 4.84; CP4.2, *eigenvalue* 1.73), con una complessità media per item (rotazione promax) di 1.2 (buona consistenza interna delle componenti individuate). La percentuale di varianza spiegata dalle due prime componenti è medio-bassa, però: di 48.43% per CP4.1 e del 17.38% per la seconda, in totale solo il 66% della varianza totale del subset di dati, da ritenersi sulla soglia dell'accettabilità (ovvero, vi è una quantità considerevole di varianza non spiegata dalle due componenti). Il CP4.1 è maggiormente rappresentato dagli item DA3,4,2,1,5 che indicano pratiche abbastanza consolidate nella didattica universitaria (dare feedback a partire dai voti ottenuti in verifiche, utilizzare i dati complessivi della valutazione del corso per costruire feedback formativi, utilizzare i voti progressivi come dati del corso per il monitoraggio della didattica, eventuale riflessione insieme agli studenti sulla composizione dei voti e progressione necessaria di performance per l'approvazione del corso). La seconda componente, non forte (17% di variabilità), ha una certa presenza della componente DA9 e DA6, con valori tuttavia limitrofi. La componente rappresenta casi relativi all'utilizzo dei dati estratti dalle piattaforme eLearning per monitorare la didattica e valutarne la qualità. Da notare che tali pratiche non implicano necessariamente l'utilizzo di strumenti avanzati come dashboards o analitiche di apprendimento predittive. La correlazione tra componenti rafforza la presenza associata della componente 4.1 e la 4.2, quindi quando vi si trova la seconda

Pratiche basate sui dati nella didattica universitaria

tipologia di pratica (più avanzata) normalmente si troverà la prima tipologia (più tradizionale). La rappresentazione grafica *biplo*t mostra una valenza positiva per tutti gli item in relazione alla prima componente (pratiche più tradizionali di assessment) mentre per la seconda componente (pratiche più avanzate), gli item DA5,3,2,1 (relativi a pratiche più tradizionali) giustamente si collocano nella zona negativa, rafforzando l'idea di una netta distinzione tra chi porta avanti pratiche di assessment basate su dati maggiormente analogici e in modalità sommative e chi adotta forme più avanzate insieme alle forme più tradizionali. Da notare un certo raggruppamento dei dati indicante differenze a livello disciplinare, in particolare tra tecnologia (con più casistica in relazione all'utilizzo di dati estratti da piattaforme LMS ed eventualmente esperienze di uso di dashboards; e l'ambito umanistico, le scienze sociali e naturali, che tendono ad orientarsi su valori negativi per entrambi le componenti. Ciò potrebbe indicare un fabbisogno formativo e riflessivo, nonché aspetti di difficoltà strutturale affrontata dai docenti per una cultura della valutazione degli apprendimenti e la didattica più spinta (l'uso dei dati è solo una modalità di approccio riflessivo a supporto di un assessment complesso e ricco) sia a livello analogico che digitale.

#### 4.2.5. SED: Potenziamento degli Studenti

Quest'ultima dimensione, come quella successiva, spostano l'attenzione dall'attività con e su dati della didattica e dei processi di apprendimento, inclusa la gestione e valutazione della qualità educativa, per spostarsi sulle competenze degli studenti volte a vivere in un mondo dove la tracciabilità dei dati può essere sia risorsa, sia pericolo.

Per quanto riguarda la scala SED, di pochi item (3) si è riscontrata la presenza di una componente principale (CP5.1, *eigenvalue* 2.04), con una complessità media per item (rotazione promax) di 1 (ottima consistenza interna della componente individuate). La percentuale di varianza spiegata dalle due prime componenti è da ritenersi media: di 68.19% per CP4.1 e del 18.54% per la seconda, quindi di per sé la prima componente spiega una percentuale accettabile di variabilità, tuttavia leggermente sotto il 70%. Il CP5.1 è rappresentato da tutti gli item, ma vede una particolare insistenza di SED1 e 2. Come nei casi precedenti, non sorprende tale risultato. Queste due componenti indicano la consistenza delle risposte relative alla condivisione di report istituzionali e di dati provenienti dalle piattaforme LMS per discutere l'avanzamento della classe o dello studente rispetto ad un programma di lavoro. Tuttavia, i valori negativi osservati nel *biplo*t per gli item relativi, consente d'indicare una bassa frequenza di queste pratiche. Il terzo item (SED3), indicante la discussione di analitiche estratte dai social media come risorsa integrata nella didattica, per una riflessione sull'avanzamento di un programma di lavoro, introduce sicuramente una pratica poco frequente. Questi risultati sottolineano che i dati sono poco condivisi con gli studenti, e se utilizzati circolano eventualmente all'interno di ambienti controllati (quelli della didattica) con meno "attraversamento" di confini istituzionali, oltre la classe, verso spazi informali come i social media. Inoltre, come ulteriore elemento d'interesse nella grafica *biplo*t si osserva il comportamento più focalizzato o centrato dell'ambito delle scienze formali positivamente su entrambi le componenti di rappresentazione, quindi a copertura di tutte le pratiche basate sui dati che mirano al potenziamento degli studenti, e con una leggera tendenza verso aspetti di

Pratiche basate sui dati nella didattica universitaria

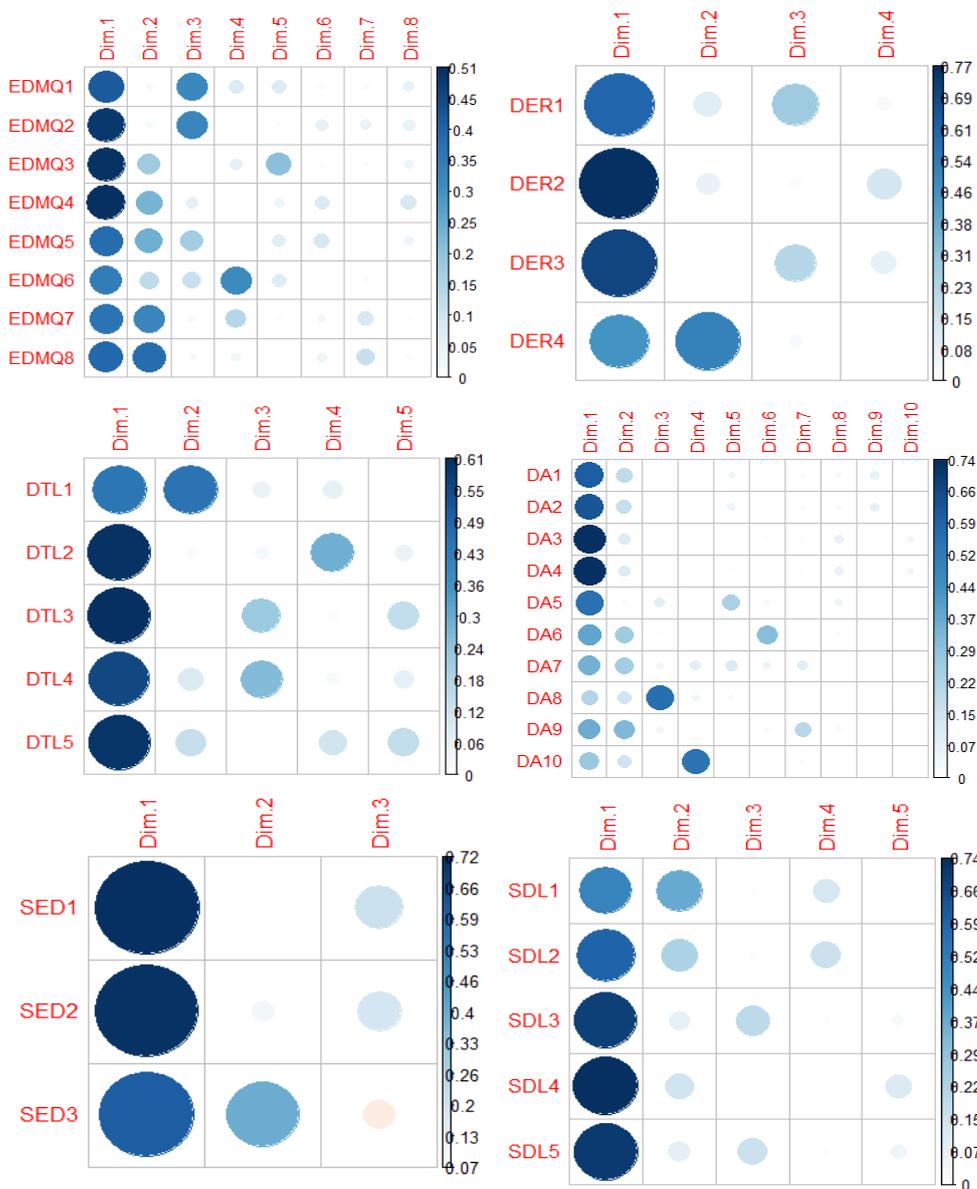
confronto con dati estratti da ambiti informali (social media, valenza più positiva dei punteggi sulla seconda componente). Gli altri ambiti disciplinari mostrano una situazione di maggiore dispersività, indicante un approccio ancora non diffuso di questa pratica basata sui dati.

#### 4.2.6. SDL: Alfabetizzazione (degli studenti) ai dati

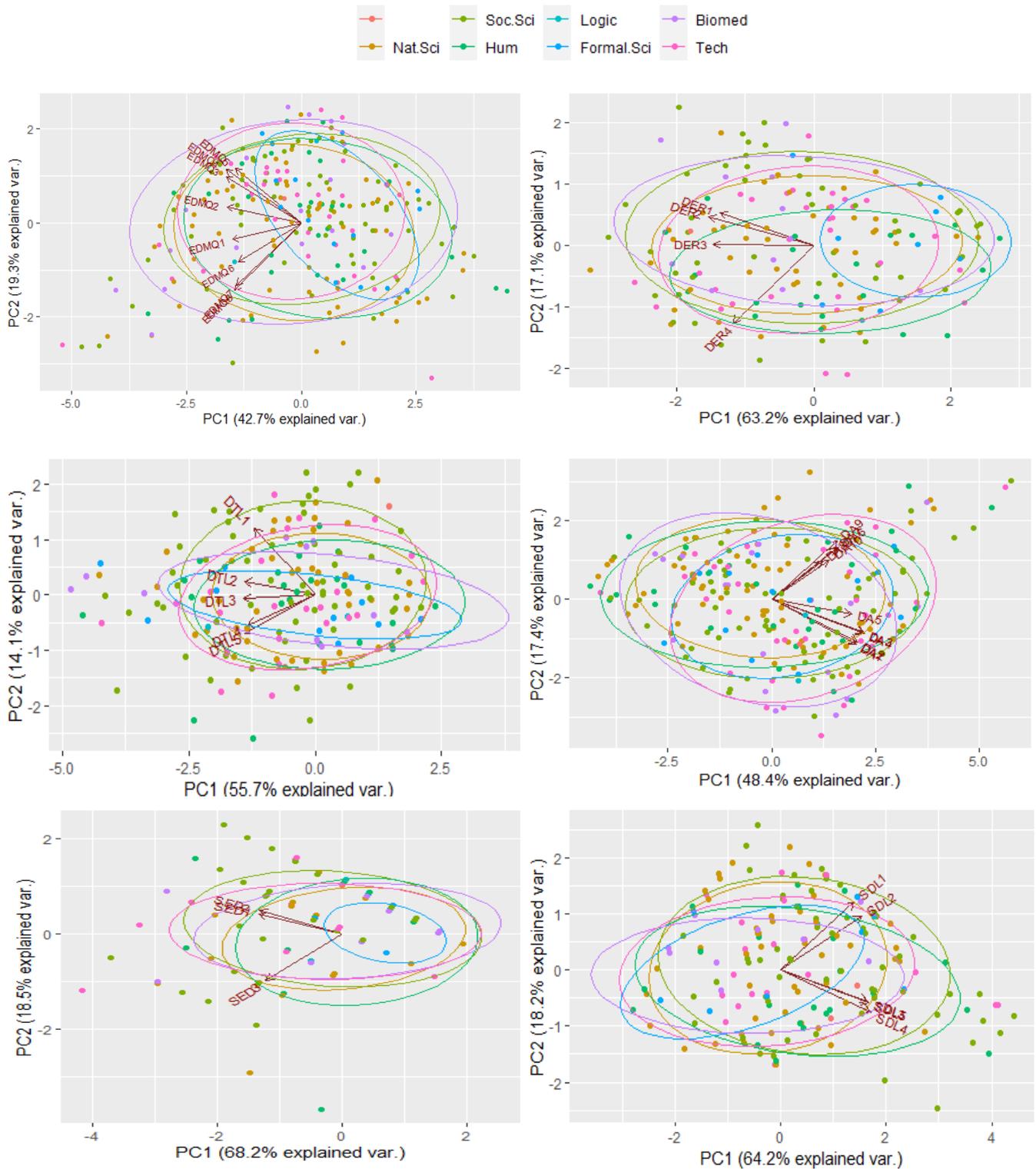
Quest'ultima dimensione, fa centraggio sul set di abilità tecniche (statistiche, computazionali e grafiche, nonché critiche) necessarie per lavorare con i dati digitali attraverso i vari ambiti professionali in cui gli studenti verranno chiamati ad operare.

La scala SDL, di 5 item ha portato all'identificazione di una evidente componente principale (CP6.1, *eigenvalue* 3.20), seguita da una seconda componente con un valore vicino alla soglia (CP6.2, 0.91). La complessità media per item (rotazione promax) si attesta su un valore di 1.2 da ritenersi buona. La percentuale di varianza spiegata dalla prima componente è del 64.19%; seguito dalla seconda (18.20%) in totale spiegano l'82.39% della variabilità. Il CP6.1 presenta una forte influenza degli item SDL3,4,5. Tale componente quindi può essere caratterizzata come una generale promozione di un approccio critico ai dati. Tale approccio, implica: la discussione e riflessione tra gli studenti sui vincoli sociali relativi all'uso dei dati (propri e altrui) e le applicazioni digitali nell'IA come processo socio-tecnico; l'analisi etico relativo raccolta e utilizzo dei propri dati come parte del processo di apprendimento; la discussione congiunta sull'impatto dell'uso dei dati nella società in generale. La seconda componente, debolmente rappresentata come abbiamo anticipato, è rappresentata soprattutto per l'item SDL1 (promozione dell'abilità degli studenti per valutare criticamente la affidabilità e credibilità dei dati incorporati in una fonte informativa). La correlazione tra la prima e la seconda componente è inferiore a quella tra la terza (troppo debole) e prima CP. Tuttavia, non può essere disdegnato il valore vicino al 50% che indica la compresenza di entrambe le componenti. La rappresentazione grafica *biplot* mostra un peso maggiore delle pratiche basate relative a SDL1 e 2 (legate ad aspetti tecnici della manipolazione dei dati e di valutazione della qualità informativa) nella prima CP. Mentre invece SDL3,4 e 5 mostrano un buon posizionamento rispetto alla prima CP (e una relativa bassa presenza sulla seconda CP. Ciò può essere letto come una integrazione di pratiche, ovvero, chi s'inoltra nell'esplorazione tecnica di rappresentazioni *data-driven*, può eventualmente includere una valorizzazione critica dei nuovi fenomeni di uso dei dati nella società e negli ambiti di riferimento professionale.

**Figura 2** – Rappresentazione grafica della contribuzione degli item per Componente/Dimensione



**Figura 3** – Rappresentazione “biplot” con integrazione degli ambiti disciplinari



## 6. Conclusioni

Alle nostre due domande di ricerca:

Pratiche basate sui dati nella didattica universitaria

- Quali sono le pratiche basate sull'uso di dati digitali nella didattica più diffuse all'interno dell'Ateneo di Padova?
- Esistono differenze attraverso le varie tipologie di pratica, considerando particolarmente l'ambito disciplinare?

Possiamo rispondere in sintesi che:

- Il campione che rispecchia la struttura tradizionale dell'università italiana. Poca presenza di donne in ambiti STEM e ai vertici. Da sospettare che le donne, negli ambiti più tradizionali, siano meno rappresentate nelle pratiche più avanzate portate avanti dalle aree STEM (in particolare DER e SDL). Tuttavia, sarebbe necessario approfondire sulla partecipazione delle donne nelle pratiche più spinte di valutazione e uso di tecniche partecipate con coinvolgimento degli studenti (come quelle riportate nelle scale SED e SDL).
- EDMQ – Prevalgono pratiche consolidate e legate all'influenza del cambiamento nazionale ed istituzionale sui sistemi di valutazione della didattica e la ricerca.
- DER – Dati come risorse educative aperte, un continente tutto da esplorare, anche se a quanto pare e per la natura stessa delle discipline, l'ambito formale (statistica, matematica, scienze computazionali) si inoltrano in questo continente. Tuttavia, da vedere che l'approccio potrebbe essere molto ricco per gli altri ambiti scientifico disciplinari.
- DTL – I dati digitali non sono ancora strutturati come risorsa al servizio della decisione docente. Non è un dato sorprendente, perché è una problematica diffusa: le piattaforme digitali generano una ingente massa di dati che non sono sufficientemente esplorati né trattati, e talvolta chi ne approfitta degli stessi sono gli stessi business legati ai servizi digitali e di *data-science*, in particolare, dietro ai social media.
- DA – Prevalgono come nell'EDMQ approcci tradizionali legati alle culture di uso di dati analogici nella valutazione. La potenzialità dei dati digitali come struttura dinamica, non viene sfruttata come elemento a supporto della discussione di feedback e formativa con gli studenti.
- SED e SDL – Una pratica ipoteticamente positiva che però non viene portata avanti sennò da chi già realizza esplorazioni tecniche dei dati. Tuttavia, si potrebbe fare una riflessione rispetto alla capacità di riflessione critica contestualizzata culturalmente senza la presenza di un dialogo interdisciplinare dove il sapere delle scienze sociali supportano una esplorazione più raffinata dei luoghi comuni attorno alle problematiche etiche suscitate dai dati. Il potenziamento e lo sviluppo di abilità di data literacy complesse ed etico-critiche (*data ethics*) è un capitolo tutto da esplorare.

## Ringraziamenti

Si ringrazia l'Ateneo di Padova per la distribuzione e ampia partecipazione nell'indagine.

Lo sviluppo del questionario e dell'approccio analitico è stato realizzato nel contesto del Progetto RYC-2016-19589 "Professional Learning Ecologies for Digital Scholarship: supporting professionalism for the modernization of Higher Education" coordinato dalla Prof.ssa Raffaghelli (UOC) ed ha contato con la collaborazione internazionale tra la UOC e l'Ateneo di Padova (Prof.ssa Valentina Grion e Prof.ssa Marina De Rossi) per l'adattamento italiano.

## References

- Cortina, J. M. (1993). What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, 78(1), 98–104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Raffaghelli, J. E. (2019). *Delphi Study to validate the survey exploring academic awareness and engagement with data-driven practices*. Barcelona. <https://doi.org/10.5281/ZENODO.3581290>
- Sangrá, A., Raffaghelli, J. E., & Guitert-Catasús, M. (2019). Learning ecologies through a lens: Ontological, methodological and applicative issues. A systematic review of the literature. *British Journal of Educational Technology*, 50(4), 1619–1638. <https://doi.org/10.1111/bjet.12795>