
Special Focus

**Il Performance management
nelle istituzioni pubbliche
dopo la pandemia da Covid-19.
Nuovi paradigmi fra riforme
e processi di cambiamento**

*Performance management in public
institutions after the Covid-19
pandemic. New paradigms between
reforms and change processes*

Valutazione della performance basata su aspettative: indagine comportamentale per l'individuazione dei bias in atto e la formulazione di possibili interventi

Expectation-based performance appraisal: behavioral investigation for the identification of current biases and formulation of possible interventions

Francesca Marzo *, Fabrizio Cafaggi **, Cecilia Colasanti ***

* Scuola Nazionale dell'Amministrazione ** Consiglio di Stato *** Istat

Abstract

Performance evaluation can become for public administrations a powerful organizational tool, also in the optic of a more pervasive adoption of work for objectives, as required by post-covid smart-working implementation. By an empirical investigation and by proposing possible behavioral interventions we aimed to contribute to this path. The empirical evidence collected confirms a correlation between assessment aimed at awarding rewards and imposing sanctions and tendency to give uniform, upwardly shifted evaluations (leniency error) and a significant difference in the importance given to different items object of evaluation. We claim that the value given to different evaluation items is due to the fact that different hierarchical levels have different systems of expectations. Reasoning about the meaning of expectations as a measurement parameter is crucial to understand what expectations are taken into account and how they are incorporated into the indicators used in evaluation systems. What emerges from the analysis is that different systems of expectations are assembled without necessary converging. One possible intervention is to highlight the whole set of expectation systems and, should they diverge, understand how this can be governed in the evaluation process. Introducing peer evaluation to flank hierarchical assessment, normally used, could be a useful debiasing tool.

Keywords: *Performance evaluation, Behavioral approach, Expectations, Bias, Leniency error.*

Introduzione

La finalità sottesa all'attività del sistema di valutazione della performance nella pubblica amministrazione non è solo quella di attribuire premi economici o favorire incentivi di carriera (Gabris & Ihrke, 2000), la cui disciplina è demandata anche alla contrattazione integrativa tra amministrazione e organizzazioni sindacali. Piuttosto, dovrebbe trattarsi di un sistema finalizzato sia all'attuazione di un insieme di verifiche e controllo dell'apporto di ciascun dipendente al raggiungimento degli obiettivi che vengono assegnati all'istituto nel suo complesso e ad ogni struttura di appartenenza in particolare, sia alla diffusione di una cultura organizzativa che vede nel raggiungimento degli obiettivi uno strumento di creazione di valore pubblico¹. Viene infatti sempre più richiesto non solo di operare per realizzare genericamente le attività che vengono assegnate dalle norme di riferimento, ma anche di raggiungere quegli specifici obiettivi assegnati ogni anno in relazione alle linee strategiche che i vari consigli di amministrazione forniscono sulla base delle indicazioni del ministro di riferimento (quindi il ministro della transizione ecologica). La valutazione, in quest'ottica, diviene strumento di carattere organizzativo, per governare le possibili differenze legate alle probabili diversità di visione sul modello organizzativo dei vari soggetti coinvolti (So *et al.*, 2015).

1. Per un'introduzione al concetto di valore pubblico e al suo legame con la valutazione delle performance si può vedere Deidda Gagliardo, E. (2015); per un approfondimento Deidda Gagliardo, E., & Saporito R. (2021).

La misurazione e la successiva valutazione della performance organizzativa e individuale sono dunque funzionali ad una consuntivazione gestionale dei risultati che viene realizzata a livello organizzativo. Misurare, che è la prima azione che compete ai dirigenti, è un atto obiettivo di per sé ed è tanto più obiettivo tanto più vengono utilizzati indicatori obiettivi. L'obiettività lascia spazio, invece, alla discrezionalità quando viene richiesto, come nel caso della performance individuale, di misurare con indicatori qualitativi. La natura degli indicatori e la loro applicazione nell'ambito del processo di valutazione possono accompagnarsi ad errori percettivi e cognitivi, specialmente quando tali processi debbano incorporare aspettative di condotta dei valutati e di soggetti terzi diversi dai valutatori. La consapevolezza della possibilità della commissione di errori, gli strumenti per la possibile correzione degli stessi consente di attribuire al processo di valutazione maggiore efficacia riducendo la percezione di discrezionalità ed arbitrarietà che ne potrebbe minare la credibilità.

Lo studio condotto e presentato nelle prossime sezioni, sviluppato a partire da un'analisi statistico-comportamentale del nuovo sistema di misurazione e valutazione della performance (da ora in poi Smvp) dell'Istituto Superiore per la Protezione e la Ricerca Ambientale (da ora in poi Ispra), adottato a partire dal 2020, al fine di approfondire alcune problematiche emerse durante la valutazione, ha portato all'individuazione di alcuni bias cognitivi dei diversi soggetti coinvolti, all'approfondimento di quelli che possono insistere sulle scelte dei valutatori e, infine, alla formulazione di alcuni interventi comportamentali sia sul breve sia sul medio periodo. Tali interventi riguardano piccole modifiche negli strumenti di rilevazione, da una parte, e interventi più importanti sull'intero processo di valutazione, dall'altra.

1. Aspetti comportamentali della valutazione

1.1. I punti di riferimento nella letteratura

Al fine di analizzare motivazioni e meccanismi relativi all'efficacia della valutazione, gli studi sviluppati nel corso degli anni sulla base di intuizioni derivanti da casi studio, si sono concentrati principalmente, da una parte, sulla progettazione dei sistemi di valutazione (indagando quali condizioni siano più favorevoli per la loro reale attuazione oppure le differenze e le somiglianze rispetto a strumenti come la gestione della qualità totale) (Bowman, 1994; Berger *et al.*, 2013; Barbieri *et al.*, 2017), dall'altra, sul tema della relazione tra i sistemi di valutazione e le percezioni dei dipendenti, le forze motivazionali e le caratteristiche demografiche (Marengo & Pasquali, 2010; Taylor & Westover, 2011; Bruno *et al.*, 2017; Coccia, 2019).

In una prospettiva comportamentale, inoltre, si è sostenuto che, fintanto che le organizzazioni vogliono che i loro sistemi di valutazione delle prestazioni raggiungano gli obiettivi prefissati, è di vitale importanza che i valutatori siano addestrati a svolgere adeguatamente

la loro funzione di valutazione e che in tale contesto la formazione comprende discussioni sui tipici errori di valutazione e strategie per limitarli (Martin & Bartol, 1986).

Allo stesso modo, Bowman (1999) ha osservato che, mentre l'uso dei rating presuppone che i valutatori siano ragionevolmente obiettivi e precisi, errori ben noti si verificano nel processo a causa di limitazioni cognitive. In aggiunta, analisi empiriche in psicologia applicata (ad esempio, Thorsteinson *et al.*, 2008) e in economia comportamentale (ad esempio, Furnham & Boo, 2011; Kahneman, 2011) suggeriscono che i punteggi delle prestazioni possono essere distorti proprio a causa di limiti cognitivi. Ancora più recentemente, infine, si è evidenziato il contributo fondamentale che un approccio comportamentale alla valutazione delle performance può permettere (Andersen & Larsen, 2016; Andersen & Hjortskov, 2016) e si è approfondito l'intervento di possibili bias nella presentazione delle informazioni riguardanti le performance dei manager della pubblica amministrazione (Belardinelli *et al.*, 2018).

Nonostante un interesse sempre crescente, si rileva una scarsità di studi empirici sui bias cognitivi dei valutatori che possono sistematicamente influenzare la loro valutazione (Belle *et al.*, 2017). Il nostro contributo muove proprio dalla necessità di contribuire a colmare questa lacuna nella comprensione dell'incidenza di errori sistematici di giudizio e delle modalità in cui tali errori cognitivi influenzano le valutazioni delle performance individuali nelle organizzazioni del settore pubblico, a partire dal caso studio riguardo il nuovo sistema di valutazione della performance dell'Ispra.

1.2. Presupposti teorici e premessa metodologica

L'approccio comportamentale alla base del nostro contributo parte da un modello che distingue due sistemi cognitivi, uno impulsivo (sistema 1) e l'altro riflessivo (sistema 2) (Kahneman, 2011) per approfondire il modo di ragionare, di decidere e di agire degli esseri umani: sin dai pionieristici studi di Simon (1979) sono stati sviluppati modelli di ragionamento umano basati su euristiche (scorciatoie del pensiero che rappresentano l'attivazione del sistema 1 a scapito del sistema 2) e i conseguenti bias cognitivi, errori sistematici indotti proprio da queste scorciatoie.

L'idea di fondo è che le nostre decisioni sono in gran parte inficiate da errori di ragionamento e quindi di giudizio. Questi difetti comportamentali producono euristiche del giudizio rapide e intuitive che si esercitano a detrimento delle nostre capacità deliberative più razionali: in base allo schema duale di Kahneman e Tversky, la nostra capacità decisionale intuitiva e impulsiva, il sistema 1, prevale sulla facoltà più razionante e riflessiva, il sistema 2.

A partire da queste considerazioni, la premessa metodologica è che la presenza di bias distorca il processo valutativo e che, quindi, una valutazione caratterizzata dall'assenza di distorsioni cognitive sia più trasparente di una influenzata da distorsioni cognitive;

l'obiettivo dell'intervento sullo strumento di rilevazione della valutazione è duplice: eliminare le fonti di bias ove possibile, mitigare gli effetti distorsivi del bias ove la eliminazione non sia possibile. Sulla base dei presupposti della premessa metodologica e delle evidenze empiriche raccolte, si potranno progettare interventi di nudging che agiscano non solo da supporto a chi progetta il sistema di valutazione, superando alcuni ostacoli, ma abilitando anche un incremento della performance individuale, incidendo sull'auto-motivazione e sui conseguenti comportamenti da adottare (Larrick, 2004; Soll *et al.*, 2014).

2. Sistemi di valutazione: riferimenti normativi, linee guida e prospettiva comportamentale

2.1. Normativa per la valutazione delle performance

Nell'ottica di un miglioramento della qualità e delle prestazioni della pubblica amministrazione, diverse novità sono state introdotte dalla normativa. Elemento distintivo e di soluzione di continuità sono la legge-delega 4 marzo 2009, n. 15 e il decreto legislativo 27 ottobre 2009, n. 150 - di seguito "Decreto" - introducono il concetto di performance organizzativa ed individuale, sostituendo definitivamente il concetto di produttività individuale e collettiva, che era stato introdotto dal decreto legislativo 3 febbraio 1993, n.29.

In particolare, il Decreto prevede che le amministrazioni pubbliche valutino annualmente la performance organizzativa e individuale e adottino, con apposito provvedimento, un sistema di misurazione e valutazione della performance. Viene inoltre stabilito che la misurazione e la valutazione della performance individuale del personale sono effettuate dai dirigenti e collegate "al raggiungimento di specifici obiettivi di gruppo o individuali, nonché alla qualità del contributo assicurato alla performance dell'unità organizzativa di appartenenza, alle competenze dimostrate ed ai comportamenti professionali e organizzativi".

Tale normativa è stata, negli anni, integrata e modificata. In particolare, facciamo riferimento al d.lgs. 74/2017 recante "Modifiche al decreto legislativo 27 ottobre 2009, n. 150, in attuazione dell'articolo 17, comma 1, lettera r), della legge 7 agosto 2015, n. 124 e alle Linee Guida n. 2/2017 emanate dal Dipartimento della funzione pubblica con l'obiettivo di fornire un supporto alle singole amministrazioni nella predisposizione dei sistemi di misurazione e valutazione della performance e nel loro miglioramento nel corso degli anni, per poter raggiungere gli obiettivi di efficacia fissati al fine di supportare i modelli organizzativi adottati. Successivamente, al fine di fornire alle amministrazioni indicazioni di maggior dettaglio in ordine alla misurazione e valutazione della performance individuale, sono state introdotte le nuove Linee Guida n. 5/2019.

A questa normativa, va ad affiancarsi, infine, il disegno di legge 80 del 9 giugno 2021 e le misure per la valorizzazione del personale e per il riconoscimento del merito in esso

previste, che di fatto non modificano quanto già previsto, ma insistono in modo più specifico sulla definizione delle classi di merito e del premio a valle del processo di valutazione.

2.2. Valutazione basata sulle aspettative e sperimentazione nelle pubbliche amministrazioni

Tra le diverse indicazioni fornite dalle nuove Linee Guida del Dipartimento della funzione pubblica, sopra menzionate, di particolare importanza per il nostro lavoro sono (a) i meccanismi di calibrazione *ex ante* ed *ex post*, (b) l'obiettivo di differenziazione delle valutazioni e (c) la centralità dei comportamenti attesi.

Il primo punto sottolinea l'importanza di metodi di valutazione verso l'obiettivo della valutazione a 360 gradi, per raggiungere la quale si può procedere alla giustapposizione di metodi di valutazione differenti alla tradizionale valutazione gerarchica (nel nostro studio, come vedremo in seguito, sono state in particolare prese in considerazione le potenzialità della valutazione tra pari). Il secondo punto, invece, evidenzia la necessità di fare in modo che la valutazione sia il più differenziata possibile, inducendo a considerare un percorso di analisi che permetta di capire quali siano, dal punto di vista cognitivo-comportamentale, gli ostacoli all'affermarsi di tale tipo di valutazione al posto della valutazione uniforme, osservabile diffusamente in molte amministrazioni. Il terzo punto, infine, ha a che fare con un importante cambiamento dell'approccio valutativo: da una valutazione basata su criteri rigidi ed astratti ad una valutazione basata sulle aspettative, che chiama in causa, per definizione, aspetti cognitivo-comportamentali, da analizzare sia a livello individuale che di interazione collettiva e organizzativa. Quest'ultimo aspetto è da considerarsi fondamentale nella comprensione delle dinamiche che portano al successo o al fallimento dei percorsi per raggiungere gli obiettivi sottesi ai primi due punti ed è stato oggetto di analisi nella lettura dell'approccio assunto da diverse amministrazioni. In particolare, sono state presi in considerazione inizialmente gli Smvp di Ispra, Anvur, Cnr, Ministero dell'ambiente, Mibact e, come ente locale, il Comune di Firenze. A queste organizzazioni sono stati aggiunti, in un secondo tempo, gli Smvp del Mit e del Mise, particolarmente innovativi e in continuo aggiornamento dal punto di vista dei criteri e dei processi, al fine di superare le difficoltà insite nella valutazione dei comportamenti stessa. Si veda, in quest'ottica, l'introduzione in entrambi questi ultimi Smvp di strategie per modificare la valutazione dei dirigenti con lo scopo di ottenere un effetto a cascata sui comportamenti dei valutatori (ad esempio, nel sistema del Mit è stato introdotto recentemente lo scarto quadratico medio per calcolare il grado di differenziazione dei giudizi in base al quale valutare la capacità dei valutatori di differenziare le valutazioni).

In base alle informazioni raccolte, è possibile affermare che, sebbene con alcune eventuali difficoltà di implementazione, diverse amministrazioni stanno puntando alla realizzazione di una valutazione dei comportamenti individuali basata sulle aspettative esplicitate in diversi descrittori. Finalità del nostro studio è stata individuare, tramite approccio comportamentale, un possibile percorso per far sì che tale realizzazione si attui in modo efficace.

2.3. Prospettiva comportamentale sui metodi di valutazione

In generale, ciascuna delle fasi del processo di misurazione e valutazione può beneficiare del contributo delle scienze comportamentali, per individuare le migliori strategie e guidare il processo decisionale di ciascun partecipante al processo (valutatore e valutato) (So *et al.*, 2015), ma anche al fine di ridurre quanto più possibile gli effetti distorsivi tipici della valutazione condotta da più valutatori (in atto a causa del fatto che, sebbene attenga il rapporto tra il singolo e il valutatore, l'intero processo si inquadra nel più ampio contesto di valutazione di tutti i soggetti che intervengono nel processo).

Una prima distinzione che può essere fatta tra i diversi metodi di valutazione riguarda la relazione tra valutato e valutatore. Nello specifico, la valutazione gerarchica è un metodo fondato sulla separazione tra valutatore e valutato, mentre nella valutazione tra pari la coincidenza tra valutatore e valutato è parziale: la dissociazione propria della valutazione gerarchica è presente, sebbene manchi la gerarchia. La comparazione tra la valutazione gerarchica e quella tra pari consente di chiarire in che modo la presenza od assenza di potere gerarchico incide sui sistemi di credenze e sui bias dei valutatori, dirigenti nel primo caso, colleghi nel secondo. Tale comparazione consente di definire le differenze e di determinare il livello di complementarità tra sistemi di valutazione (Cafaggi, 2020).

Nei sistemi basati sulle aspettative, oggetto di valutazione è la conformità o la difformità tra la performance e le aspettative del valutatore, tenendo presente che, alle aspettative del valutatore, vanno aggiunte quelle dell'organizzazione che valuta i valutatori e quelle degli stessi valutati: l'esito del percorso valutativo è determinato dalla combinazione tra i diversi sistemi di aspettative. Tuttavia, anche nelle ipotesi in cui le schede di valutazione non facciano esplicito riferimento alle aspettative, queste svolgono un ruolo fondamentale nel processo valutativo e costituiscono pilastro essenziale dell'architettura comportamentale (Cafaggi, 2021).

Tra i metodi di valutazione si è scelto di indagare la valutazione tra pari ed il rapporto tra questa e quella gerarchica, con particolare riguardo alle differenze esistenti a seconda che vi sia o no un rapporto di gerarchia tra valutatore e valutato e se dal processo di valutazione derivino conseguenze legate all'attribuzione di premi o l'irrogazione di sanzioni. La ricerca sui profili comportamentali ha aiutato ad indagare non solo in che modo possono distinguersi i metodi di valutazione dei pari e di valutazione gerarchica ma anche in che modo i bias dei soggetti coinvolti nel processo mutano e quali correttivi possono essere apportati al fine di mitigare le distorsioni cognitive. In particolare, l'esame comparativo tra valutazione gerarchica e valutazione tra pari ha consentito di definire in che modo la seconda possa aiutare la soluzione di distorsioni cognitive relative alla prima. In particolare, tramite l'indagine empirica si è voluto indagare due profili: la relazione tra le distorsioni cognitive di coloro che hanno predisposto le schede e la distorsione cognitiva di coloro che hanno proceduto alla valutazione; in questo secondo ambito, le possibili differenze tra valutazione gerarchica e valutazione tra pari, nonché le possibili influenze

che la valutazione dei dirigenti subisce per il fatto che i dirigenti stessi sono a loro volta soggetti a valutazione.

3. Indagine empirica per validare l'analisi comportamentale

3.1. Questionario per indagare il leniency error

Come si è detto nell'introduzione, l'indagine è stata sviluppata a partire da un'analisi preliminare dei dati raccolti dal sistema Smvp di Ispra, in base alla quale è stata osservata la tendenza a dare valutazioni omogenee, spostate verso l'alto (problema che si riscontra per tutti gli item e per tutte le valutazioni raccolte), rendendo ipotizzabile l'intervento di un errore di distribuzione noto in letteratura come leniency error (Kane *et al.*, 1995; Yildiz & Baltaci, 2009; Ng *et al.*, 2011).

Obiettivo della ricerca empirica è stato dunque indagare le caratteristiche di questo fenomeno nel contesto considerato, cioè all'interno dell'istituto Ispra, raccogliendo dati sulle valutazioni effettuate dai singoli soggetti chiamati a valutare nel sistema Smvp adottato. In particolare, ci si è focalizzati sull'eventuale diversa incidenza del fenomeno nei casi in cui le valutazioni hanno dirette conseguenze sulla irrogazione di sanzioni e l'assegnazione di premi e nei casi in cui le valutazioni hanno solo l'obiettivo di acquisire informazioni, senza che ne derivino conseguenze dirette sul piano economico.

Le ipotesi alla base di questo studio sono le seguenti: la prima è che in un ambiente in cui le valutazioni servono esplicitamente a creare premi/punizioni si dovrebbe osservare un più forte leniency error rispetto ad un ambiente in cui la valutazione opera solo come strumento per il miglioramento delle performance; la seconda è che in un ambiente in cui la valutazione gerarchica è esplicitamente utilizzata per distribuire premi/punizioni e quella tra pari solo come strumento di calibrazione, si dovrebbe osservare un leniency error maggiore nella valutazione gerarchica rispetto alla valutazione tra pari; la terza (social proof) è che quando il network di riferimento tende a dare dei giudizi uniformi il valutatore (sia pari che gerarchico) tenderà ad uniformare le proprie valutazioni.

L'indagine empirica si è avvalsa di un questionario, somministrato via google form al personale dirigente e non dirigente di tre diverse amministrazioni (due amministrazioni centrali e un ente locale).

Al fine di cogliere la differenza tra valutazioni finalizzate all'irrogazione di sanzioni e l'assegnazione di premi e valutazioni con il solo obiettivo di acquisire informazioni senza che ne derivino conseguenze dirette il questionario è stato sviluppato immaginando due scenari:

- al rispondente viene chiesto di immedesimarsi nel ruolo di dirigente;

- al rispondente viene chiesto di immedesimarsi nel ruolo di collega.

Un secondo e correlato profilo è quello relativo alle conseguenze della valutazione: la correlazione con premi e sanzioni pregiudica l'obiettività della valutazione orientandola ed in particolare impedendo che vi sia un'adeguata differenziazione tra i valutati? Più specificamente si è testata l'ipotesi che una valutazione priva di conseguenze possa produrre una maggiore differenziazione e dunque esalti il merito rispetto ad una valutazione cui siano associate conseguenze immediate ed in particolare l'assegnazione di premi e l'irrogazione di sanzioni.

Integrando queste osservazioni nei due scenari sopra riportati, si possono ottenere le seguenti possibilità: che le valutazioni dei dirigenti abbiano conseguenze economiche dirette mentre quelle dei pari no; che sia le valutazioni dei dirigenti sia quelle dei pari incidano sull'assegnazione di premi e l'irrogazione di sanzioni; né le valutazioni dei dirigenti né quelle dei pari abbiano conseguenze perché il valore della valutazione è legato ai sistemi motivazionali e non agli incentivi economici.

Un terzo profilo d'indagine concerne il valore delle norme sociali (Bicchieri, 2017). Si assume che vi sia una correlazione tra le modalità di valutazione e le norme sociali presenti nella organizzazione. Si vuole quindi indagare se le norme sociali influenzano la valutazione e come lo fanno. In particolare, due norme sociali: quella concernente il ruolo della gerarchia. In che modo viene percepita la differenza tra superiore gerarchico e collega nel processo di valutazione; quella concernente il conformismo ai valori dell'organizzazione e dunque al rapporto tra merito che dovrebbe risultare nella differenziazione dei giudizi e egualitarismo che esita nell' assenza di difformità.

3.2. Elaborazione dei dati raccolti, differenze e sovrapposizioni nella valutazione gerarchica ed in quella tra pari

È necessario premettere che, se il numero dei soggetti rispondenti appartenenti al personale non dirigente (3204) fornisce una rappresentazione statisticamente significativa (convergenza dei risultati), il numero dei soggetti appartenenti all'area dirigenziale (61) non fornisce una statistica altrettanto completa. Ciononostante, alcune tendenze rilevate nelle risposte fornite dagli uni e dagli altri, permettono di identificare similitudini e differenze significative nelle priorità delle due categorie.

In particolare, dai dati riportati di seguito nel grafico in Fig. 1, si evince ad esempio che le competenze e il lavoro di squadra sono ritenute fondamentali da entrambe le categorie. Circa il 36% dei dirigenti e il 38% del personale non dirigente sceglie la voce "competenze e conoscenze" come prima tra quelle proposte. Il lavoro di squadra è invece scelto come primo rispettivamente dal 21% dei dirigenti e dal 25% del personale non dirigente.

Fig. 1 Confronto tra le prime scelte (barre) ed i totali pesati riportati al numero dei soggetti (MP) per dirigenti (light gray) e personale non dirigente (dark gray) (fonte: elaborazione degli autori su dati raccolti).

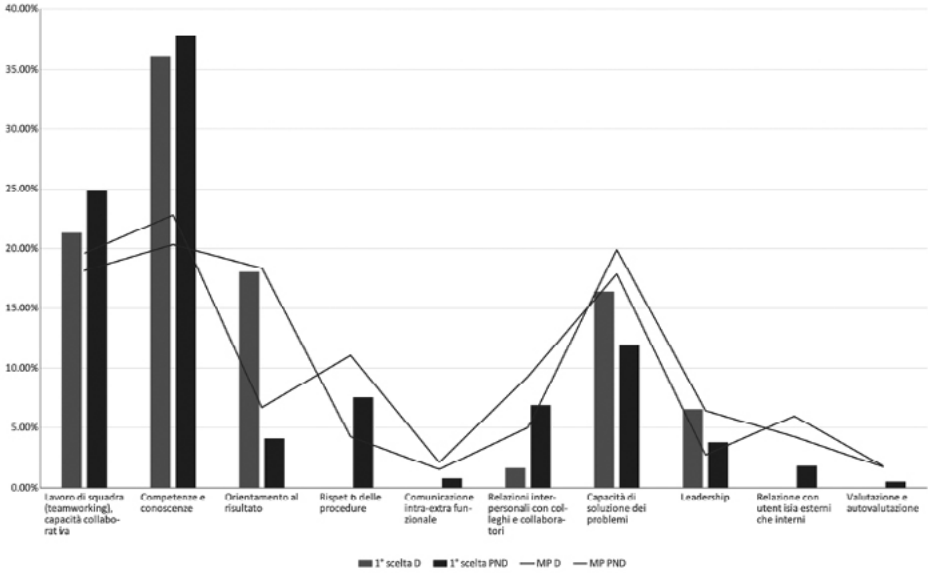
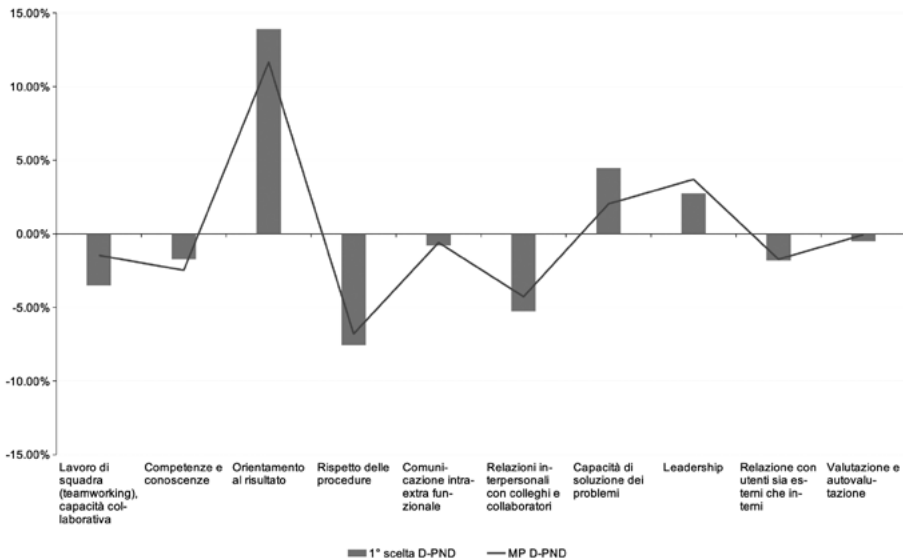


Fig. 2 Differenze algebriche tra le prime scelte ed i totali pesati riportati al numero dei soggetti (MP) tra dirigenti e personale non dirigente (fonte: elaborazione degli autori su dati raccolti).



La differenza sostanziale tra le due categorie si verifica tra le voci “orientamento al risultato” preferito dal 18% dei dirigenti e soltanto dal 4% del personale non dirigente. Discorso invertito per quanto riguarda il “rispetto delle procedure”, scelto come primo da circa l’8% del personale non dirigente e da nessun dirigente. La differenza tra le scelte è evidenziata dall’ovale dove, insieme alle prime scelte, sono riportati i totali pesati tra le cinque opzioni e riportati al numero dei soggetti (Mp) di dirigenti (grigio chiaro) e personale non dirigente (grigio scuro). Altra differenza, meno importante numericamente ma comunque significativa riguarda le voci “relazioni interpersonali con colleghi e collaboratori” preferita dal personale non dirigente e “capacità di soluzione dei problemi” maggiormente presente nei riscontri relativi ai dirigenti. Entrambe le differenze sottolineano come il personale dirigente sia maggiormente focalizzato sull’obiettivo, mentre il personale non dirigente dia molta importanza alle modalità lavorative. La situazione è evidenziata dal grafico in Fig. 2 dove sono rappresentate le differenze tra prime scelte e totali pesati riportati al numero dei soggetti (media pesata, Mp) tra personale dirigente (D) e non dirigente (Pnd). Qui, i valori positivi significano che la scelta è preferita dai dirigenti (rettangolo superiore), mentre i valori negativi corrispondono alle scelte preferite dal personale non dirigente (rettangolo inferiore).

Per quanto riguarda le risposte relative alla preferenza tra valutazione differenziata e uniforme, quelle dei dirigenti e del personale non dirigente si equivalgono (rispettivamente 88% e 87% a favore della differenziata). Inoltre, nonostante la netta maggioranza di dirigenti e personale non dirigente si ritenga a favore della valutazione differenziata, la percentuale di chi crede che gli altri siano in maggioranza a favore di una valutazione uniforme è maggiore di chi crede il contrario. In particolare, sia il personale dirigente che il personale non dirigente crede mediamente che più della metà dei parigrado effettuerebbe una valutazione uniforme, rispettivamente 56% e 55%.

Fig. 3 Credenza riguardo la percentuale di parigrado che farebbe valutazioni uniformi (fonte: elaborazione degli autori su dati raccolti).

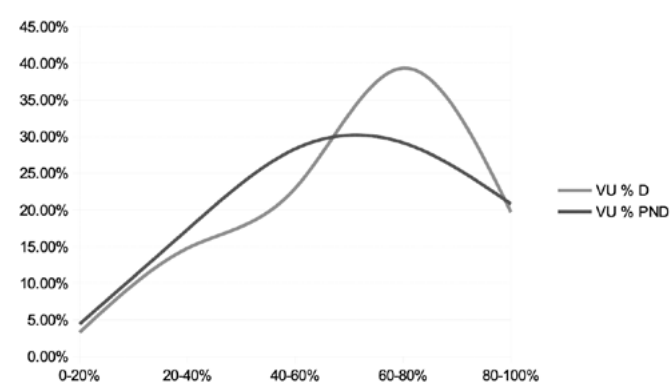
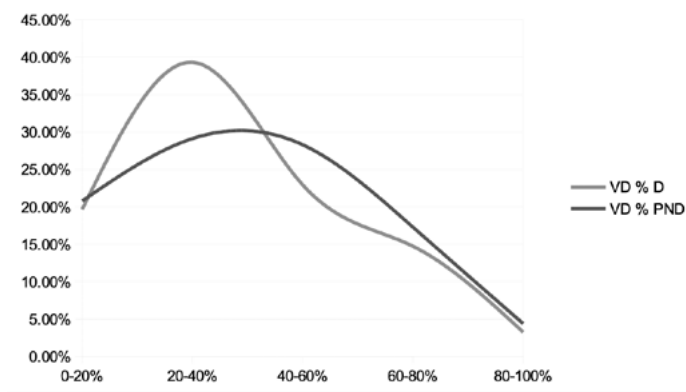


Fig. 4 Credenza riguardo la percentuale di parigrado che farebbe valutazioni differenziate (fonte: elaborazione degli autori su dati raccolti).



In Fig. 3 è rappresentata la distribuzione in frequenza per dirigenti e personale non dirigente delle credenze relative al tipo di valutazione che ci si aspetta che effettuerebbero i parigrado. Le curve sono relative alla valutazione uniforme. Le curve relative alla valutazione differenziata sono simmetriche rispetto a quelle della valutazione uniforme (Fig. 4).

3.3. Analisi delle correlazioni tra domande nei due questionari

Per il questionario sottoposto ai dirigenti, prima di tutto, si è proceduto con la correlazione per indagare la relazione tra le preferenze per lo spazio discrezionale e quelle relative al tipo di valutazione (uniforme vs differenziata). L'ipotesi è che la preferenza verso le valutazioni uniformi/differenziate varia a seconda dell'attitudine verso lo spazio discrezionale consentito nel valutare (se lo si ritiene adeguato, se ne desidera di più o di meno). Un test chi-quadro di Pearson effettuato sulla cross-tabulazione tra le due variabili evidenzia una correlazione statisticamente significativa ($p\text{-value} = 0.027$) tra le due variabili.

In particolare, andando ad analizzare i coefficienti delle singole categorie tramite un modello di regressione logistica, questa relazione sembra resistere solo per quanto riguarda chi preferirebbe uno spazio discrezionale per la valutazione minore. Confrontando la probabilità di scegliere una valutazione differenziata piuttosto che uniforme di chi preferirebbe uno spazio discrezionale maggiore/minore con quella di chi lo ritiene adeguato, risulta che preferire un minor spazio discrezionale corrisponde ad una minor probabilità di differenziare la valutazione rispetto al ritenerlo adeguato, sebbene questo non significhi che ad una preferenza per uno spazio maggiore di discrezionalità corrisponda una maggior probabilità di differenziare la valutazione. Al contrario, chi ritiene lo spazio discrezionale adeguato e chi vorrebbe vederlo aumentato tende ad avere la stessa probabilità di procedere con una valutazione differenziata.

Per quanto riguarda il questionario sottoposto al personale non dirigente, si è deciso di analizzare innanzitutto la correlazione tra la percezione di efficacia dei benefici pecuniari e gli aspetti che si ritiene siano influenzati dalla valutazione, a partire dall'ipotesi che chi ritiene più efficaci i benefici pecuniari riterrà che la valutazione influenzi principalmente gli item più "materialistici" (e.g. retribuzione, raggiungimento dei risultati), mentre chi non ritiene efficaci i benefici pecuniari riterrà che la valutazione influenzi principalmente gli item meno "materialistici" e più legati alla qualità del lavoro (e.g. conferma dei valori personali, relazioni tra colleghi). Un test chi-quadro di Pearson effettuato sulla cross-tabulazione tra le due variabili evidenzia una correlazione statisticamente significativa ($p\text{-value} = 0.001$) tra la variabile "Efficacia dei benefici pecuniari" e ciascuna delle tre variabili indicanti rispettivamente il primo, il secondo ed il terzo item indicati ed ordinati per importanza.

Visto l'alto numero di categorie delle variabili indicanti i tre aspetti, per interpretarne meglio il significato e la rilevanza, è stato sviluppato un modello di regressione logistica multinomiale per poter analizzare i coefficienti delle singole categorie. Per ogni iterazione del modello, vengono stimati coefficienti che indicano come ritenere efficaci i benefici pecuniari aumenti o diminuisca la probabilità di selezionare ciascuno degli item sulla lista rispetto alla probabilità di selezionare un item scelto come categoria di riferimento. Il modello è stato fatto girare 10 volte, scegliendo ogni volta una categoria di riferimento diversa (tranne la categoria "Altro") per poter analizzare tutte le possibili combinazioni. Per ogni iterazione, genere, età, livello di istruzione, livello professionale e tipo di amministrazione di provenienza sono stati inseriti come variabili di controllo.

Dall'analisi risulta che ritenere efficaci i benefici pecuniari:

- Aumenta la probabilità di indicare "retribuzione" come prima scelta rispetto a tutte le altre opzioni;
- Aumenta la probabilità di indicare "raggiungimento dei risultati" come prima scelta rispetto alle opzioni "conferma dei valori personali", "lavoro di squadra (team-working), capacità collaborative", "qualità del lavoro", "relazione con i dirigenti" e "relazione tra colleghi";
- Aumenta la probabilità di indicare "qualità del lavoro" come prima scelta rispetto all'opzione "relazione tra colleghi";
- Aumenta la probabilità di indicare "riconoscimento dell'impegno" come prima scelta rispetto all'opzione "relazione tra colleghi".

I risultati sembrano rafforzare le aspettative previste dall'ipotesi: per chi ritiene efficaci i benefici pecuniari, la probabilità di ritenere che la valutazione influenzi in primo luogo

la retribuzione aumenta (prevedibilmente) rispetto a tutte le altre possibili opzioni. Per quanto riguarda altri item considerabili come più “materialistici”, come “raggiungimento dei risultati” e “riconoscimento dell’impegno”, la probabilità di essere indicati come l’aspetto più influenzato aumenta solo rispetto ad alcune categorie, le quali possono essere considerate come meno “materialistiche”.

A seguire, si è effettuata l’analisi per indagare l’eventuale correlazione tra la percezione di efficacia dei benefici pecuniari e la credenza normativa, in base all’ipotesi che chi ritiene efficaci i benefici pecuniari riterrà più giusto differenziare la valutazione, e viceversa. Un test chi-quadro di Pearson effettuato sulla cross-tabulazione tra le due variabili evidenzia una correlazione statisticamente significativa ($p\text{-value} = 0.001$) tra le due variabili. Sebbene entrambi i gruppi (chi ritiene efficaci i benefici pecuniari e chi non li ritiene efficaci) presentino una maggioranza in favore delle valutazioni differenziate, la categoria “valutazione differenziata” risulta positivamente associata con il ritenere efficaci i benefici pecuniari.

Per confermare la rilevanza della differenza tra i due gruppi, è stato effettuato un test d’ipotesi sulla significatività statistica della differenza tra i due gruppi. La proporzione di chi preferisce la valutazione differenziata alla valutazione uniforme varia di 6,70 punti percentuali tra chi ritiene efficaci i benefici pecuniari e chi non li ritiene efficaci. Il $p\text{-value}$ associato al test d’ipotesi su questa differenza (0.0001) permette di rifiutare l’ipotesi nulla che non ci sia differenza tra i due gruppi e concludere che questa differenza è statisticamente significativa. Ne risulta, dunque, che chi ritiene efficaci i benefici pecuniari tende a ritenere più giusto differenziare la valutazione, e viceversa. La differenza della percentuale di chi ritiene giusto differenziare tra i due gruppi è stimata al 6,70%, con un intervallo di confidenza al 95% di (4,00% - 9,40%).

Si è svolta poi l’analisi per indagare quale relazione è osservabile tra le aspettative riguardo le conseguenze di una valutazione differenziata in un contesto in cui si tende ad uniformare le valutazioni e l’aderenza o meno alle norme di comportamento. L’ipotesi di base è che chi si aspetta conseguenze negative per l’aver differenziato la valutazione quando gli altri uniformano tenderà a voler differenziare di meno quando gli altri uniformano, e viceversa. Un test chi-quadro di Pearson effettuato sulla cross-tabulazione tra le due variabili evidenzia una correlazione statisticamente significativa ($p\text{-value} = 0.001$) tra le due variabili. Sebbene in entrambi i gruppi (chi ritiene che valutare in maniera differenziata porti a conseguenze e chi non lo ritiene) presentino una maggioranza in favore delle valutazioni differenziate, la categoria “Valutazione Differenziata” risulta positivamente associata con il ritenere che valutare in maniera differenziata quando gli altri uniformano porti a delle conseguenze per il valutatore. Per tentare di comprendere meglio questo risultato contrario alle aspettative, lo stesso test è stato ripetuto tra la variabile “Se è diffusa una valutazione uniforme, differenzi?” e la variabile indicante le conseguenze indicate dai vari soggetti che hanno risposto di aspettarsi conseguenze al quesito 17. Anche in questo caso, un test chi-quadro di Pearson evidenzia una correlazione statisticamente

significativa ($p\text{-value} = 0.001$) tra le due variabili. Per interpretare meglio il significato e la rilevanza delle varie categorie della variabile indicante le conseguenze attese, si procede con un modello di regressione logit per poter analizzare i coefficienti delle singole categorie. Anche controllando per genere, età, livello di istruzione, livello professionale e tipo di amministrazione di provenienza, tutti i coefficienti delle singole categorie (tranne la categoria residua “altro”) risultano statisticamente significativi. I coefficienti (calcolati adottando come categoria di riferimento il non aver dato nessuna risposta, e quindi il non aspettarsi conseguenze) sembrano essere in linea con il risultato precedente, avendo tutti un segno positivo tranne il coefficiente per la categoria “sfiducia”, che è negativo. Questo significa che a tutte le possibili conseguenze attese corrisponde una probabilità maggiore di procedere comunque con una valutazione differenziata rispetto alla probabilità associata con il non aspettarsi conseguenze (tranne per la conseguenza “sfiducia, a cui corrisponde una minor probabilità).

Dall’analisi effettuata si può notare che i risultati sono contrari all’ipotesi iniziale: chi si aspetta conseguenze tende a differenziare la valutazione più spesso di chi non se ne aspetta. Andando ad analizzare le specifiche conseguenze attese, osserviamo che l’ipotesi è confermata solo nei confronti di chi si aspetta “sfiducia” (una conseguenza negativa, che porta ad una minor propensione a differenziare) e “vantaggi reputazionali” (una conseguenza positiva, che porta ad una maggior propensione a differenziare). In tutti gli altri casi in cui sono state indicate conseguenze, si tratta di conseguenze negative che però portano in ogni caso ad una maggior propensione a differenziare la valutazione.

A seguire, si è analizzata la correlazione tra gli elementi importanti per la valutazione e le credenze normative, in base all’ipotesi che a diverse preferenze indicate riguardo gli item più importanti per la valutazione corrispondano diverse probabilità di ritenere adeguata la valutazione differenziata piuttosto che quella uniforme. Un test chi-quadro di Pearson effettuato sulla cross-tabulazione tra le due variabili evidenzia una correlazione statisticamente significativa ($p\text{-value} = 0.001$; 0.034 e 0.05 rispettivamente) tra la variabile indicante le credenze normative sulla preferenza tra valutazione uniforme e differenziata e le prime tre variabili indicanti rispettivamente il primo, il secondo ed il terzo item indicati ed ordinati per importanza. Al contrario, le correlazioni con le due variabili indicanti il quarto ed il quinto item per importanza non risultano statisticamente significative ($p\text{-value} = 0.099$ e 0.863). Per interpretare meglio il significato e la rilevanza delle varie categorie della variabile indicante le preferenze, è stato sviluppato un modello di regressione logistica per poter analizzare i coefficienti delle singole categorie. Per ogni iterazione del modello, vengono stimati coefficienti che indicano come selezionare ciascuno degli item sulla lista aumenti o diminuisca la probabilità di ritenere adeguata la valutazione differenziata rispetto alla probabilità di ritenere adeguata la valutazione uniforme rispetto a chi ha indicato l’item scelto come categoria di riferimento. Il modello è stato fatto girare 10 volte, scegliendo ogni volta una categoria di riferimento diversa per poter analizzare tutte le possibili combinazioni. Per ogni iterazione, genere, età, livello di istruzione, livello

professionale, tipo di amministrazione di provenienza e le preferenze indicate al secondo ed al terzo posto sono stati inseriti come variabili di controllo. Le preferenze indicate al quarto e quinto posto non sono state incluse per la scarsa rilevanza statistica individuata precedentemente.

Dall'analisi risulta che:

- Indicare come prima scelta “competenze e conoscenze” aumenta la probabilità di ritenere adeguata la valutazione differenziata rispetto alla probabilità di ritenere adeguata la valutazione uniforme, rispetto a chi ha indicato come prima scelta “capacità di soluzione dei problemi”, “rispetto delle procedure”, “relazioni interpersonali con colleghi e collaboratori”, “lavoro di squadra (team-working), capacità collaborative” e “relazione con utenti sia esterni che interni”;
- Indicare come prima scelta “leadership” aumenta la probabilità di ritenere adeguata la valutazione differenziata rispetto alla probabilità di ritenere adeguata la valutazione uniforme, rispetto a chi ha indicato come prima scelta “capacità di soluzione dei problemi”, “rispetto delle procedure”, “relazioni interpersonali con colleghi e collaboratori”, “lavoro di squadra (team-working), capacità collaborative” e “relazione con utenti sia esterni che interni”;
- Indicare come prima scelta “orientamento al risultato” aumenta la probabilità di ritenere adeguata la valutazione differenziata rispetto alla probabilità di ritenere adeguata la valutazione uniforme, rispetto a chi ha indicato come prima scelta “relazioni interpersonali con colleghi e collaboratori” e “relazione con utenti sia esterni che interni”.

Per quanto riguarda le variabili di controllo, genere ed educazione risultano essere statisticamente significativi: gli uomini risultano meno inclini a differenziare rispetto alle donne, mentre il livello di istruzione è positivamente correlato alla propensione a valutare in maniera differenziata.

L'analisi rileva delle associazioni statisticamente significative tra alcune preferenze espresse per gli item di valutazione e la propensione a ritenere adeguata la valutazione differenziata; tuttavia, si tratta di differenze tra specifiche categorie e non generali (nessuna preferenza aumenta la probabilità di differenziare rispetto a tutte le altre preferenze).

Infine, si è voluto analizzare la correlazione tra le risposte al quesito sugli elementi importanti per la valutazione, le aspettative riguardo l'impatto sulla collaboratività dei colleghi e le risposte al quesito sulle credenze normative, riproponendo la correlazione incentrata sulla risposta “lavoro di squadra (team-working)”, “capacità collaborative” inserendo “collaboratività dei colleghi di fronte ad una valutazione differenziata” come controllo.

Chi indica come preferenza valori come il lavoro di squadra (team-working), infatti, potrebbe avere un atteggiamento diverso nei confronti della valutazione se ritiene che una valutazione differenziata abbia un effetto positivo piuttosto che negativo sulla collaboratività dei colleghi. Il modello di regressione è simile a quello impiegato per la correlazione precedente, ma è stato usato per stimare l'effetto dello scegliere il lavoro di squadra rispetto a qualsiasi altra scelta e controllando per la risposta al quesito sull'impatto sulla collaboratività.

Il primo modello, che stima solo gli effetti principali delle due variabili, riporta una correlazione negativa significativa ($p < 0.001$) tra l'indicare il lavoro di squadra come prima scelta ed il preferire le valutazioni differenziate, ed una correlazione positiva tra il ritenere che la valutazione differenziata aumenti la collaboratività dei colleghi ed il preferire le valutazioni differenziate (questa relazione stranamente non regge per la categoria "molto più collaborativo", per la quale non si evidenziano differenze significative rispetto alla categoria centrale di riferimento).

Andando ad analizzare il modello con le interazioni, si evidenzia che l'effetto della variabile "lavoro di squadra" varia, in effetti, a seconda dell'opinione sull'impatto della valutazione differenziata sulla collaboratività dei colleghi. La correlazione negativa vista nel modello main effects, infatti, regge solo per le categorie "meno collaborativo", "più collaborativo" e "molto più collaborativo". Per chi ritiene che i colleghi risultino "molto meno collaborativi" o "altrettanto collaborativi",² invece, non risultano effetti significativi.

Dall'analisi risulta che il ritenere che la valutazione differenziata aumenti la collaboratività dei colleghi risulta positivamente correlato con il preferire le valutazioni differenziate, come da ipotesi. Tuttavia, l'effetto dell'item "lavoro di squadra" è più inatteso: indicarlo come prima scelta non ha nessun effetto sulla tendenza a preferire la valutazione differenziata per le persone che ritengono che quest'ultima non influenzi la collaboratività dei colleghi, mentre mostra una correlazione negativa sia per chi ritiene che la collaboratività dei colleghi aumenti, sia per chi ritiene che la collaboratività dei colleghi diminuisca.

4. Interpretazioni comportamentali e implicazioni operative per la gestione dei sistemi di valutazione

Tra le diverse interpretazioni possibili dei risultati fin qui delineati, è possibile delinearne una prettamente comportamentale, finalizzati alla modifica di alcuni aspetti dei sistemi di misurazione e valutazione delle performance, che sia in grado, nel tempo di sollecitare e supportare l'instaurarsi di comportamenti virtuosi riguardo la differenziazione delle

2. La non significatività di questo gruppo potrebbe essere dovuta al margine d'errore più ampio dovuto a minor numero di osservazioni in questa categoria, piuttosto che ad una vera assenza di un effetto.

valutazioni, nello specifico, e l'atteggiamento nei confronti della valutazione delle performance all'interno delle organizzazioni, più in generale.

Precisando che l'implementazione di tali interventi necessita di ulteriori indagini empiriche per testarne l'efficacia e la sostenibilità, qui di seguito viene descritta la prospettiva comportamentale e gli interventi proposti.

4.1. Incentivi, leniency error e aspettative

Per quanto riguarda i risultati della survey, ciò che si può inferire è che la persistenza delle valutazioni uniformi sia attribuibile all'esistenza di una diretta causalità tra valutazione effettuata e attribuzione di premi o irrogazione di sanzioni. In particolare, le correlazioni osservate indicano che, quando c'è un'alta considerazione dell'efficacia degli incentivi economici, si alza la probabilità di considerare come più importanti le ripercussioni della valutazione sugli aspetti più "materialistici" rispetto a quelli relazionali. Da ciò si può dedurre che gli incentivi economici producono una sorta di "frame" sull'utilità di un sistema di valutazione che taglia fuori gli aspetti di collaborazione e di clima lavorativo (relazione tra colleghi, teamworking, etc.) e concentra l'attenzione sui risultati più tangibili, quali la retribuzione e il raggiungimento dei risultati.

Di conseguenza, non percependo di fatto alcuna ripercussione potenzialmente positiva sugli aspetti più qualitativi del lavoro svolto di una valutazione effettuata in modo corretto e, quindi, differenziato, non solo ci si concentrerà sulla falsa credenza che una valutazione differenziata serva solo ad attribuire premi e irrogare sanzioni, ma anche sul fatto che essa potrà solo produrre effetti negativi sulla collaboratività. Questa percezione distorta potrebbe essere la causa che induce i soggetti, che pure si esprimono positivamente nei confronti della correttezza di una valutazione differenziata e che nello scenario simulato scelgono proprio di differenziare, ad un'inversione di preferenza verso una valutazione uniforme quando si tratta di valutare nel mondo reale.

Da sottolineare la differenza di impatto in generale che si considera come conseguenza della valutazione: i dirigenti tendono a considerarlo consistente, mentre per i pari grado non si percepisce di intensità elevata. Questo fatto porta a credere che tale inversione di preferenza abbia luogo in particolare proprio tra i dirigenti, che, tra l'altro, sono coloro la cui valutazione andrà effettivamente a produrre ripercussioni tangibili dal punto di vista degli incentivi economici. Tuttavia, per quanto riguarda la seconda ipotesi generale alla base dello studio sviluppato tramite il questionario, non si osservano le differenze attese nella tipologia di valutazione scelta nello scenario concreto proposto: sia i dirigenti che i pari grado scelgono di differenziare.

Un ultimo punto da evidenziare a proposito dei risultati della survey è che è possibile, data l'osservazione di una diversità di pesi attribuiti ai diversi item di valutazione tra personale

dirigente e valutatori pari grado, confermare che la riduzione della discrezionalità nei criteri di valutazione non elimina di fatto la necessità di un'esplicitazione delle aspettative per ridurre il rischio di arbitrarietà nella valutazione. Se si pensa che, poi, bassi spazi di discrezionalità possono anche ostacolare eventuali strategie di riduzione del fenomeno della valutazione uniforme, tale esplicitazione diventa quasi un'urgenza.

4.2. Valutazione tra pari come strumento di debiasing

Come abbiamo visto, le evidenze empiriche confermano una correlazione tra valutazione finalizzata alla attribuzione di premi e all'irrogazione di sanzioni e la tendenza a dare valutazioni uniformi e spostate verso l'alto (leniency error). Analogamente, i risultati dell'indagine svolta confermano una differenza significativa nell'ordine di importanza che viene dato ai diversi item su cui la valutazione dovrebbe concentrarsi.

Collegando questi risultati, si possono fare alcune ipotesi per spiegare i comportamenti osservati e proporre alcune considerazioni generali per fare luce sulle questioni sollevate dalle domande fondamentali alla base della ricerca e, in particolare, su come favorire la percezione della valutazione come mezzo/percorso per il miglioramento organizzativo piuttosto che come strumento di controllo. In particolare, si può immaginare che il valore che viene dato a diversi item di valutazione sia collegato ad aspettative diverse alla base dell'intero sistema di valutazione e che, quindi, la differenza osservata tra le preferenze dei dirigenti e quelle dei pari grado sia un segnale della diversità delle reciproche aspettative.

In questa prospettiva, il punto cruciale è che i livelli gerarchici possono avere visioni e sistemi di aspettative diversi e che è importante fare emergere queste aspettative per governare e non occultare la diversità, al fine di attuare un processo di acquisizione di consapevolezza. Ragionare sul significato delle aspettative come parametro di misurazione è fondamentale per capire esattamente quali aspettative vengono prese in considerazione e come esse vengono incorporate negli indicatori utilizzati nei sistemi di valutazione.

Quello che emerge dall'analisi è che convergono, nella valutazione fatta dai dirigenti rispetto ai collaboratori e ai dipendenti, almeno tre sistemi di aspettative che incidono sulla valutazione:

- le aspettative di chi redige la scheda di valutazione (della dirigenza) rispetto a chi valuta;
- le aspettative dei valutatori rispetto ai valutatori, ovvero il rapporto tra i diversi dirigenti, cioè le aspettative che ci si aspetta gli altri dirigenti abbiano rispetto alla propria unità;
- le aspettative dei valutati.

Ciononostante, ciò che viene esplicitamente evidenziato nelle schede di valutazione difficilmente ha a che fare con altro che l'aspettativa del valutatore rispetto al comportamento del valutato: gli altri due sistemi, pur giocando un ruolo fondamentale, non emergono e rimangono solo endogenamente concorrenti. Una strategia di intervento possibile è quella di mettere in rilievo tutto l'insieme dei sistemi di aspettative e, nel caso di loro divergenza, comprendere in che modo essa possa essere governata nel processo di valutazione.

Con questa finalità, l'introduzione della valutazione tra pari ad affiancare quella gerarchica, normalmente utilizzata in tutti i sistemi di valutazione, può essere considerato un utile strumento.

Infatti, la comparazione tra la valutazione gerarchica e quella tra pari consente di chiarire in che modo la presenza od assenza di potere gerarchico incide sui sistemi di credenze e sui bias dei valutatori, dirigenti nel primo caso, colleghi nel secondo.

Tale comparazione permette di definire le differenze e di determinare il livello di complementarità tra sistemi di valutazione. I sistemi motivazionali del valutatore gerarchico possono differire da quelli impiegati dai pari nel valutare la performance dei propri colleghi dando luogo a risultati almeno parzialmente differenti. In quest'ottica la possibilità di condurre entrambi i metodi di valutazione in contemporanea può costituire una tattica comportamentale per esplicitare le differenze tra aspettative, da una parte, e avere uno strumento per pesare i risultati del processo di valutazione, dall'altra.

Un intervento di questo tipo, oltre a contribuire dunque ad un'esplicitazione di eventuali conflitti, in grado di rendere trasparente il confronto e di non relegarlo solo a materia di mediazioni esterne, può aiutare anche nel superamento del lenency error. Richiamando, nell'esplicitazione dei sistemi di aspettative differenti, l'obiettivo di riconduzione alla finalità di miglioramento organizzativo tutto il sistema di valutazione si possono infatti raggiungere due obiettivi propedeutici a questo superamento: l'allontanamento dalla prospettiva di premi e punizioni che, abbiamo visto, amplifica la motivazione di uniformare le valutazioni e la possibilità di non dover dimostrare obiettività ma di essere apertamente portatori di una determinata prospettiva all'interno di un sistema complesso di ruoli di valutatori e valutati. Ovviamente, tutto questo potrà realizzarsi solo se sorretto da un adeguato sistema di norme sociali su cui occorre parimenti lavorare.

4.3. Il ruolo delle norme sociali nella valutazione di performance

Il profilo d'indagine concerne il valore delle norme sociali, per il quale abbiamo ipotizzato un legame tra le modalità di valutazione e le norme sociali presenti nella organizzazione, ci ha portato a formulare nella survey alcune ipotesi riguardo l'influenza delle norme sociali sulla valutazione delle performance individuali e riguardo il come quest'influenza si può esplicitare.

In questo quadro, ricordiamo che un obiettivo fondamentale dell'indagine empirica è stato quello di comprendere l'esistenza di un eventuale conformismo ai valori dell'organizzazione e dunque al rapporto tra merito che dovrebbe risultare nella differenziazione dei giudizi e egualitarismo che esita nell'assenza di difformità.

Per quanto riguarda l'influenza delle norme sociali sulla valutazione, i dati raccolti mostrano che, quando i valutatori sono chiamati ad un ragionamento deontico, le credenze normative personali sia dei dirigenti sia dei pari grado sono in linea con le linee guida e si orientano verso la differenziazione della valutazione. Quando invece si passa alle credenze normative di tipo empirico (cosa mi aspetto che gli altri facciano), si osserva un interessante ribaltamento: i soggetti, sia dirigenti sia pari grado, si aspettano che la maggior parte degli altri valutatori propenderà per una valutazione uniforme.

Si può notare una sorta di sfiducia nei confronti di ciò che le norme sociali esistenti prescrivono: pur considerando corretto differenziare, mi aspetto che la tendenza generale sia di uniformare la valutazione. Una considerazione di questo tipo potrebbe essere la causa di un "intention-action gap" che fa sì che la maggioranza dei soggetti intervistati sostiene che si comporterebbe come la norma deontica prescrive (differenziare la valutazione in caso di generalizzata uniformità) e invece, nella realtà, si conforma alla norma sociale empirica.

Tra le cause di questo comportamento, come mostrano i dati raccolti, vi è l'aspettativa di concrete conseguenze di uno scostamento da tali norme, identificate sia dai dirigenti sia dai pari soprattutto in possibili conflitti nelle relazioni lavorative.

Infine, osservando le correlazioni tra alcuni quesiti chiave e quelli prettamente formulati per indagare l'influenza delle norme sociali, si possono estrapolare alcune regolarità interessanti. In primo luogo, il fatto che chi ritiene efficaci i benefici pecuniari tende a ritenere più giusto differenziare la valutazione riporta la nostra attenzione a quanto sostenuto nei paragrafi precedenti riguardo la percezione della valutazione come strumento per attribuire premi e irrogare sanzioni: in questo caso la norma deontica di differenziazione viene motivata dalla finalità stessa che si percepisce dell'intero sistema di valutazione. Quando invece ci si muove verso le aspettative empiriche, si osserva che, a differenza di quanto si potrebbe ipotizzare, chi si aspetta ripercussioni per il fatto di differenziare le proprie valutazioni quando gli altri le uniformano, tende a scegliere la differenziazione come norma di comportamento corretta. La spiegazione di questo atteggiamento è da ricercarsi nel fatto che, probabilmente, in una cultura organizzativa che si percepisce come giudicante e pronta a "punire", si è pronti a propendere per un comportamento che ci differenzi e ci porti nella direzione di "fare la cosa giusta", nonostante non sia la più conveniente.

Un ulteriore dato risultante dalle correlazioni è quello che pone in relazione le credenze normative deontiche e le preferenze riguardo gli item su cui è importante valutare le performance individuali: l'analisi, come si è visto, rileva delle associazioni statisticamente

significative tra il considerare giusta la valutazione differenziata e la scelta di item più competitivi e orientati al risultato (quali proprio “orientamento al risultato”, “competenze e conoscenze” e “leadership”). Anche in questo caso, questa correlazione mostra un legame forte tra preferenza per la valutazione differenziata e la finalità del sistema di valutazione più legato agli incentivi.

Procedendo alla “correlazione di controllo”, mettendo in una relazione tripla le risposte al quesito riguardante le preferenze sugli item più orientate alla collaborazione, le risposte riguardo l’impatto della valutazione sulla collaboratività e quelle sulle credenze normative, si può ipotizzare che chi esprime preferenza tra gli item sul lavoro di squadra, avrà un atteggiamento diverso nei confronti della valutazione se ritiene che la valutazione differenziata abbia un effetto positivo sulla collaboratività. Tale ipotesi risulta confermata dall’analisi effettuata, sebbene, in maniera inattesa, la scelta del “lavoro di squadra” come item più importante ha un effetto sulla tendenza a preferire la valutazione differenziata solo per chi crede che essa avrà impatto sulla collaboratività dei colleghi: più precisamente, se si crede che la valutazione influenzi la cooperatività, non importa se si crede che la aumenti o la diminuisca, si tenderà a scegliere di meno la valutazione differenziata come deonticamente giusta.

In un quadro complesso come quello evidenziato da questi risultati, appare chiara l’importanza di un lavoro dedicato ad interventi comportamentali mirati ad intervenire su queste percezioni, per poter trovare strategie di norm-nudging in grado di supportare la differenziazione delle valutazioni intervenendo sulle dinamiche alla base della formazione delle aspettative empiriche.

Riflessioni conclusive

L’analisi comportamentale ha contribuito ad evidenziare alcune debolezze del processo valutativo esaminando il rapporto tra organizzazione, valutatori e valutati e le relative aspettative, tenendo presenti tre diversi livelli di analisi: chi valuta, come si valuta, cosa si valuta. Cercando di comprendere le relazioni tra i diversi livelli, si è deciso di mirare ad incentivare una valutazione del merito differenziata e disincentivare un modello di valutazione che esita nella distribuzione di incentivi uguali per tutti.

Per quanto riguarda la valutazione della performance individuale, uno degli aspetti innovativi dei nuovi Smvp è il riflettore posto su quelli che sono definiti i “comportamenti organizzativi” richiesti per il più efficace svolgimento delle funzioni assegnate, richiamando l’attenzione non solo sui risultati, ma anche sul “come” questi risultati si raggiungono.

L’enfasi sugli aspetti comportamentali viene in particolare sottolineata dalla scelta del metodo plurifattoriale per la valutazione della prestazione, fondato, nello specifico, sulle

competenze professionali dimostrate (comportamenti organizzativi) e la performance organizzativa dell'unità di appartenenza. Considerando che i comportamenti possono essere interpretati come prodotti dalle competenze agite dalla persona, diventano il segnalatore della disponibilità e dell'uso delle competenze stesse. In questa visione del modello teorico misto di valutazione, il concetto di performance diventa più ampio e abbraccia, insieme ai risultati/obiettivi (più o meno quantificabili), i comportamenti messi in atto per realizzarli. Tra gli elementi chiave di processo e di contenuto utilizzati per identificare la tipologia di valutazione da utilizzare, infatti, due punti in particolare delle Linee guida del Dipartimento della funzione pubblica vanno in questa direzione: l'enfasi sui processi organizzativi e non solo sui risultati, in modo tale da poter valutare gli outcome (impatti) anziché gli output (risultati), poiché la producibilità degli impatti attesi di lungo termine è possibile proprio grazie al controllo dei processi organizzativi, dei comportamenti e delle qualità delle scelte effettuate; il coinvolgimento delle risorse umane, con l'intento di cogliere, attraverso l'implementazione del modello, un'occasione di formazione individuale e di creazione di nuove convinzioni e valori manageriali nei gruppi.

Nello studio effettuato, ci si è concentrati sui valutatori e si prospetta il coinvolgimento degli individui chiamati a valutare anche in fase di implementazione e monitoraggio degli Svmp, permettendo e sollecitando confronti sulle diverse aspettative e sulle prospettive riguardo cosa valutare, per avviare un processo di architettura concertata delle scelte relative alla performance e agli item da valutare. In altri termini, si è avviato un processo che mira ad esplicitare le aspettative degli attori coinvolti nella valutazione, permettendo allo stesso tempo di determinare in che misura le performance individuali sono autonome ed in che misura dipendono da aspetti organizzativi, considerando che tale dipendenza può avere caratteristiche non solo motivazionali, ma anche strutturali.

La prospettiva futura riguarda un'implementazione sperimentale da effettuare in alcune amministrazioni tra quelle prese in esame, apportando alcune modifiche sia agli strumenti utilizzati sia all'intero processo di valutazione e monitorando l'efficacia di tali interventi comportamentali effettuati parallelamente ad un percorso formativo riguardante la valutazione di performance.

Si tratta dunque di un processo che necessita di essere proseguito per allineare gli obiettivi dell'organizzazione con quelli dei valutatori grazie alla possibilità di far emergere le divergenze di credenze esistenti in modo trasparente e consapevole. In altri termini, l'idea è di proseguire con l'approccio comportamentale, affiancando le diverse strategie di intervento più strutturale che sono state individuate, per fare in modo che la riflessione su questi aspetti diventi momento di formazione e spinta verso possibili cambiamenti culturali. L'obiettivo è di progettare interventi di norm-nudging, a partire dalla misurazione e dall'analisi delle attuali aspettative, misurazione e analisi che devono necessariamente proseguire in modo concertato, attraverso focus group formativi e laboratori sistematici.

Un ulteriore sviluppo della ricerca riguarda l'approfondimento della prospettiva dei valutati, fondamentale, al pari di quella dei valutatori sui quali ci si è focalizzati nel presente studio. A partire da una distinzione tra processo decisionale del valutatore e processo decisionale del valutato, si potrà mirare ad utilizzare l'approccio comportamentale per indagare empiricamente gli aspetti motivazionali che intervengono nel relazionarsi, da parte dei singoli individui all'interno dell'organizzazione, con la valutazione, le sue potenzialità e le sue conseguenze. Grazie a questa nuova prospettiva sarà inoltre possibile investigare come l'interazione di una pluralità di componenti cognitive, relazionali e di contesto spinga l'individuo a mettere in atto un determinato comportamento non solo nei confronti del sistema di valutazione, ma anche del proprio ruolo nello svolgimento delle attività lavorative.

Il fine è quello di integrare, poi, la prospettiva sui valutatori e quella sui valutati, per progettare e implementare strategie comportamentali di supporto ad una governance collaborativa, che implichi la condivisione degli obiettivi nella fase della programmazione (i.e. definizione e comunicazione delle aspettative), e la partecipazione alla definizione del sistema di valutazione, in un percorso di apprendimento continuo (i.e. fasi successive e di feedback).

Riferimenti Bibliografici

- Andersen, S.C., & Hjortskov, M. (2016). Cognitive Biases in Performance Evaluations. *Journal of Public Administration Research and Theory*, 26(4).
- Antoni, C.H., Baeten, X., Perkins, S.J., Shaw, J.D., & Vartiainen, M. (2017). Reward Management: Linking Employee Motivation and Organizational Performance. *Journal of Personnel Psychology*, 16(2).
- Baard, P. P., Deci, E. L., & Ryan, R. M. (2004). Intrinsic need satisfaction: A motivational basis of performance and well-being in two work settings. *Journal of Applied Social Psychology*, 34(10).
- Barbieri, M., Girosante, G., & Valotti, G. (2017). *Comparative Analysis of the Performance Evaluation Systems of Public Sector Employees*. Milano: Egea Edizioni.
- Battaglio, R.P., Belardinelli, P., Belle, N., & Cantarelli, P. (2018). Behavioral Public Administration ad fontes: A Synthesis of Research on Bounded Rationality, Cognitive Biases, and Nudging in Public Organizations. *Public Administration Review*, 79(3).
- Belardinelli, P., Belle, N., Sicilia, M., & Steccolini, I. (2018). Framing Effects under Different Uses of Performance Information: An Experimental Study on Public Managers. *Public Administration Review*, 78(6).
- Belle, N., Cantarelli, P., & Belardinelli, P. (2017). Cognitive Biases in Performance Appraisal: Experimental Evidence on Anchoring and Halo Effects with Public Sector Managers and Employees. *Review of Public Personnel Administration*, 37(3).
- Berger, E.M., Hrabring, C., & Sliwka, D. (2013). Performance Appraisals and the impact of forced distribution – An experimental investigation. *Management Science*, 59(1).
- Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford: Oxford University Press.
- Bowman, J. S. (1994). At last, an alternative to performance appraisal: Total quality management. *Public Administration Review*, 54.
- Bowman, J. S. (1999). Performance appraisal: Verisimilitude trumps veracity. *Public Personnel Management*, 28.
- Brown, R. W. (1982). Performance appraisal: A policy implementation analysis. *Review of Public Personnel Administration*, 2.
- Bruno, B., Faggini, M., & Parziale, A. (2017). Motivation, Incentives and Performance: An Interdisciplinary Review. *International Journal of Business and Management*, 12(12).
- Castelfranchi, C. & Lorini, E. (2003). Cognitive Anatomy and Functions of Expectations, in Schmalhofer, F., Young, R., & Katz, G. (Eds.), *Proceedings of Eurocogsci 03. The European Cognitive Science Conference 2003*. London: Routledge.
- Cantarelli, P., Belle, N., & Belardinelli, P. (2020). Behavioral public HR: Experimental evidence on cognitive biases and debiasing interventions. *Review of Public Personnel Administration*, 40(1).
- Chinotti, O., & Festa, A. (2008). *Nuovi strumenti per lo sviluppo delle persone nelle organizzazioni: la tecnica della valutazione a 360 °*. Milano: Franco Angeli.
- Coccia, M. (2019). Intrinsic and extrinsic incentives to support motivation and performance of public organizations. *Journal of Economics Bibliography*, 6(1).
- Cohen, M., & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *The Journal of neuroscience*, 27(2).

- Deidda Gagliardo, E. (2015). *Il Valore Pubblico, la nuova frontiera delle performance*. Roma: RIREA.
- Deidda Gagliardo, E. & Saporito, R. (2021). Il Piao come strumento di programmazione integrata per la creazione di valore pubblico. *Rivista Italiana di Public Management*, 4(2).
- Dipartimento della funzione pubblica (2017). Linee guida per il Sistema di misurazione e valutazione della performance dei ministeri, 2.
- Dipartimento della funzione pubblica (2019). Linee guida per la misurazione e valutazione della performance individuale, 5.
- European Commission. Directorate-General for Employment, Social Affairs and Inclusion (2015). *Quality of Public Administration. Toolbox for Practitioners*.
- Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics*, 40.
- Gabris, G. T., & Ihrke, D. M. (2000). Improving employee acceptance toward performance appraisal and merit pay systems: The role of leadership credibility. *Review of Public Personnel Administration*, 20.
- Kahneman, D. (2011). *Thinking, fast and slow*. New York City: Macmillan.
- Kane, J. S., Bernardin, S. H., Villanova, P., & Peyrefitte, J. (1995). Stability Of Rater Leniency: Three Studies. *Academy of Management Journal*, 38(4).
- Larrick, R. P. (2004). Debiasing, in Powsner, R. A., & Powsner, E. R. (Eds.), *Blackwell handbook of judgment and decision making*.
- Martin, D. C., & Bartol, K. M. (1986). Training the raters: A key to effective performance appraisal. *Public Personnel Management*, 15.
- Marengo, L., & Pasquali, C. (2010). How to get what you want when you do not know what you want: A model of incentives, organizational structure and learning. *LEM Working Paper Series*, 8.
- Ng, K. Y., Koh, C., Ang, S., Kennedy, J. C., & Chan, K. Y. (2011). Rating leniency and halo in multisource feedback ratings: testing cultural assumptions of power distance and individualism-collectivism. *Journal of Applied Psychology*, 96.
- O’ Doherty, J.P. (2004). Reward representations and reward-related learning in the human brain: insights from neuroimaging. *Current Opinion in Neurobiology*, 14.
- OECD (2017). *Engaging Public Employees for a High-Performing Civil Service*, Organization for Economic.
- Oliva, G., Setola, R., & Scala, A. (2019). Sparse analytic hierarchy process: an experimental analysis. *Soft Computing*, 23.
- Ostinelli, G. (2005). *Motivazione e comportamento: le variabili psicologiche necessarie per raggiungere obiettivi*. Trento: Erickson.
- Rowe, J.B., Eckstein, D., Braver, T., & Owen, A. M. (2008). How Does Reward Expectation Influence Cognition in the Human Brain? *Journal of Cognitive Neuroscience*, 20(11).
- So, J., Chethana, A., DaHee, H. Nidhi, A., Adam, D., & Durairaj, M. (2015). The Psychology of Appraisal: Specific Emotions and Decision-Making. *Journal of Consumer Psychology*, 25(10).
- Soll, J.B., Milkman, K. L., & Payne, J. W. (2014). A User’s Guide to Debiasing, In Keren, G., & Wu, G. (Eds.), *Handbook of Judgment and Decision Making*. Wiley-Blackwell.
- Thaler, R. H., & Sunstein, C. R. (2018). *La spinta gentile. La nuova strategia per migliorare le nostre decisioni su denaro, salute, felicità*. Milano: Feltrinelli.
- Thorsteinson, T. J., Breier, J., Atwell, A., Hamilton, C., & Privette, M. (2008). Anchoring effects on

performance judgments. *Organizational Behavior and Human Decision Processes*, 107.

- Taylor, J., & Westover, J. H. (2011). Job satisfaction in the public service: The effects of public service motivation, workplace attributes and work relations. *Public Management Review*, 13.
- Van Eerde, W., & Thierry, H. (1996). Vroom's expectancy models and work-related criteria: A meta-analysis. *Journal of Applied Psychology*, 81(5).
- Viale, R. (2018). *Oltre il nudge. Libertà di scelta, felicità e comportamento*. Bologna: Il Mulino.
- Vroom, V.H. 1964. *Work and motivation*. New York City: Wiley.
- Yıldız, G., & Baltacı, A. (2009). Leniency and Severity Errors in Performance Appraisal in the Context of Collectivist and Individualist Culture. International Symposium on Sustainable Development, Saraybosna, Bosna-Hersek.