

Tra oggettività, discrezionalità e standardizzazione: analisi storico-critica delle rubriche ministeriali del colloquio all'esame di maturità

Cristiano Corsini*, Conny De Vincenzo*, Beatrice Partouche**

Premessa

Il colloquio dell'esame di Stato rappresenta una preziosa lente per osservare le trasformazioni nella valutazione scolastica italiana. A differenza dello scritto, più facilmente riconducibile a criteri di correzione formalizzati, la prova orale richiede una prestazione complessa, situata e interattiva, nella quale interagiscono conoscenze disciplinari, capacità argomentative, uso del linguaggio, raccordi tra saperi e modalità di relazione con la commissione. Proprio per questa ragione, la valutazione del colloquio rappresenta potenzialmente uno spazio ad alta discrezionalità, nel quale il giudizio è spesso affidato a criteri impliciti, consuetudini professionali, rappresentazioni della maturità personale e mediazioni interne alle commissioni. Dal 2020, l'introduzione di rubriche nazionali ha segnato un tentativo di rafforzare l'affidabilità della valutazione dell'orale.

Il presente contributo, storico-critico con analisi documentale, muove da una ricostruzione dell'evoluzione della valutazione del colloquio nell'esame conclusivo della scuola secondaria, con particolare attenzione al rapporto tra discrezionalità, criteri espliciti e standardizzazione. Il lavoro si basa su

Rec. 8/5/2026 app. 21/05/2026

* Cristiano Corsini, Conny De Vincenzo, Dipartimento di Scienze della formazione, Università Roma Tre; cristiano.corsini@uniroma3.it; conny.devincenzo@uniroma3.it.

** Beatrice Partouche, Dipartimento di Scienze umane, economiche e giuridiche, Università Europea di Roma; beatrice.partouche@unier.it.

Il contributo è il frutto di un lavoro congiunto. In particolare, C. Corsini ha redatto la premessa, le conclusioni e il paragrafo 6; C. De Vincenzo i paragrafi 4 e 5; B. Partouche i paragrafi 1, 2 e 3.

Cadmo (ISSNe 1972-5019), 2026, 1, DOI: 10.3280/cad2026oa23297

registri d'esame, relazioni di commissione, normativa primaria e secondaria, rubriche ministeriali dal 2020 al 2026. In particolare, l'analisi dei dispositivi ministeriali si incentra sulle categorie di criterio, indicatore, descrittore, livello, punteggio ed evidenze osservabili. L'obiettivo è quello di valutare criticamente le rubriche recentemente introdotte, chiedendosi se esse contribuiscano effettivamente a migliorare la validità e l'affidabilità dei giudizi.

1. Dalle prassi locali all'esame di Stato

La questione dei criteri di valutazione dell'esame finale della scuola secondaria superiore percorre l'intera storia del sistema scolastico italiano, intrecciandosi con le successive ridefinizioni delle finalità degli studi e con il dibattito docimologico sulla soggettività della valutazione (Morandi, 2025). La ricerca docimologica, avviata nei primi anni Trenta da Henri Piéron attraverso l'analisi degli esami di baccalauréat francese, aveva già documentato la forte soggettività intrinseca alle prove tradizionali e la grande diversità dei criteri di giudizio adottati dai commissari di fronte alle medesime prestazioni (De Landsheere, 1971; Piéron, 1963).

Nel contesto italiano, questa problematica trova riscontro già nella fase pre-gentiliana: l'analisi dei registri d'esame dell'epoca documenta come l'assenza di parametri condivisi si traducesse in soluzioni valutative non uniformi. Se i registri d'esame del regio liceo "Terenzio Mamiani" di Roma (1908-1914) mostrano una rilevazione valutativa per ciascuna prova¹, i registri coevi delle scuole normali applicano modelli di quantificazione aggregata le cui scale numeriche variano arbitrariamente da sede a sede². La coesistenza di questi modelli documenta come la valutazione conclusiva variasse in base alle consuetudini amministrative dei singoli istituti.

Con la riforma Gentile fu istituito l'esame di maturità per i licei: un impianto centralizzato e uniforme che tuttavia non si tradusse in criteri espliciti di valutazione³. La novità strutturale più rilevante sul piano valutativo risiedeva nell'articolo 92, che imponeva il principio della completa estraneità dei valutatori: "ciascun commissario deve astenersi dal giudicare i candidati che

¹ Archivio storico del liceo ginnasio statale "Terenzio Mamiani" di Roma, Registri degli esami di licenza ginnasiale e liceale, anni scolastici 1907/1908-1913/1914.

² ACS, MPI, Direzione generale Istruzione classica scientifica e magistrale, Registri degli esami di licenza e abilitazione nelle scuole normali, anni scolastici 1913/1914-1914/1915 (sedi di Rossano Calabro, Rovigo, Reggio Calabria e altre).

³ Regio decreto 6 maggio 1923, n. 1054, "Ordinamento dell'istruzione media e dei convitti nazionali".

siano stati da esso privatamente o pubblicamente istruiti”. Le commissioni, costituite interamente da membri esterni, esaminavano gli studenti fuori dalle proprie scuole d’origine su programmi estesi all’intero triennio liceale, senza conoscere il percorso formativo dei candidati e dunque chiamate a formulare un giudizio basato esclusivamente sulla prestazione della seduta d’esame (Cives, 1994).

I registri della maturità classica del liceo Mamiani di Roma relativi alle sessioni 1924-1927 documentano questa struttura⁴: la qualifica conclusiva “maturo” o “non maturo” veniva apposta nella colonna delle osservazioni, senza che alcun criterio esplicito ne motivasse l’attribuzione.

2. La maturità come giudizio sulla persona

La crescita demografica e le riforme strutturali degli anni Sessanta determinarono la transizione verso la scuola di massa. In risposta alle istanze di trasformazione del sistema scolastico la riforma Sullo⁵ introdusse una struttura d’esame sperimentale e provvisoria, che tuttavia rimase in vigore per un trentennio (Galfré, 2020).

L’articolo 5 ridefinì la finalità stessa dell’esame, orientandolo alla “valutazione globale della personalità del candidato considerata con riguardo anche ai suoi orientamenti culturali e professionali”, accompagnata da un “motivato giudizio” basato sull’esame e sul curriculum degli studi. Per la prima volta un testo normativo nominava esplicitamente il criterio valutativo del colloquio - la “personalità” del candidato - senza tuttavia definire parametri per la sua misurazione.

La relazione degli esami di abilitazione magistrale nella sede di Velletri per l’anno scolastico 1967-1968, redatta prima che la riforma Sullo entrasse in vigore, documenta i criteri adottati nella conduzione del colloquio: il presidente annotava che “la condotta delle prove orali fu sempre rivolta più alla ricerca della personalità complessiva e alla maturità specifica dei candidati, che non alla provvisoria preparazione”, dando “più peso alla fisionomia intellettuale e morale del candidato che al suo occasionale bagaglio di nozioni”⁶.

⁴ Archivio storico del “Liceo ginnasio statale Terenzio Mamiani” di Roma, Registri degli esami di maturità classica, sessioni 1924/1925-1926/1927.

⁵ Decreto legislativo 15 febbraio 1969, n. 9, convertito nella legge 5 aprile 1969, n. 119, “Riordinamento degli esami di Stato di maturità, di abilitazione e di licenza della scuola media”.

⁶ ACS, MPI, Direzione generale Istruzione classica, Relazione del presidente della prima commissione giudicatrice per gli esami di abilitazione magistrale nella sede di Velletri “D. Falconi”, anno scolastico 1967/1968.

Questi verbali attestano come la valutazione del colloquio si fondasse, prima ancora della riforma Sullo, su giudizi complessivi di tipo personologico, relativi alla maturità, alla preparazione culturale generale, all'orientamento, senza riferimento a criteri condivisi o verificabili, in piena continuità con quanto De Landsheere (1971) aveva identificato come i principali fattori di variabilità soggettiva nella valutazione orale.

Le relazioni degli anni successivi alla riforma confermano questo orientamento. Gli esami di maturità classica presso il liceo "Orazio" di Roma per l'anno scolastico 1974-1975 attestano colloqui condotti "prescindendo dal vecchio sistema di domanda e risposta, effettuando una vera discussione sugli argomenti svolti con inquadramenti nel tempo storico e in parallelo, in divergenza o in convergenza con altri periodi", utilizzando i contenuti del programma ministeriale come "piattaforma" o "punto di partenza" per avviare "discorsi ad ampio raggio" e dialoghi su argomenti "intesi come latamente culturali", orientati a rilevare le "effettive capacità intellettive e la sensibilità degli esaminandi"⁷.

Il registro della maturità classica dell'istituto "Marcantonio Colonna" di Roma per la sessione del 1969, nella sezione denominata "Giudizio dell'esame e valutazione sull'orientamento", consente di osservare questa impostazione applicata ai singoli candidati⁸. I giudizi descrivevano tratti caratteriali e attitudinali: "elemento non sempre riflessivo e impegnato al massimo, rivela normali capacità di apprendimento" (39/60); "elemento che si distingue per la serietà e la costante applicazione, rivela una preparazione e un livello culturale superiore alla media" (43/60), mentre le valutazioni sull'orientamento indicavano percorsi universitari consigliati o sconsigliati: "si sconsigliano studi di medicina e scientifici"; "il candidato non palesa particolari capacità di attendere a superiori corsi di studi". I verbali attestano inoltre come i punteggi venissero frequentemente rimodulati in sede di scrutinio finale in funzione di valutazioni complessive sulla persona piuttosto che sulla singola prestazione⁹.

⁷ ACS, MPI, Direzione generale Istruzione classica, Relazione finale relativa agli esami di maturità classica sostenuti presso il liceo classico "Orazio" di Roma, anno scolastico 1974-1975.

⁸ Archivio storico dell'istituto "Marcantonio Colonna" di Roma, Registro degli esami di maturità classica, III commissione, anno scolastico 1968-1969, verbale di valutazione sull'orientamento.

⁹ ACS, MPI, Direzione generale Istruzione classica, Relazione generale sugli esami svolti dalla I commissione esami di maturità classica operante presso il liceo classico "T. Tasso" di Roma, anno scolastico 1974-1975.

3. Dalla discrezionalità alla standardizzazione

Il modello della riforma Sullo venne sostituito dalla legge 10 dicembre 1997, n. 425, regolata dal DPR 23 luglio 1998, n. 323, che ridefinì la denominazione dell'esame conclusivo in "esame di Stato" a partire dalla sessione 1998/99. L'impianto normativo introdusse il credito scolastico triennale, tre prove scritte e un colloquio orale su "argomenti di interesse pluridisciplinare attinenti ai programmi e al lavoro didattico dell'ultimo anno di corso"¹⁰. Il punteggio fu espresso in centesimi, con soglia di sufficienza a 60, e la commissione divenne mista e paritaria. Neanche in questa occasione, tuttavia, furono introdotti criteri espliciti di valutazione del colloquio.

La prima introduzione di criteri formalizzati di valutazione riguardò le prove scritte. Il decreto legislativo 13 aprile 2017, n. 62, emanato in attuazione della legge n. 107 del 2015, stabilì all'articolo 17 che le "griglie"¹¹ di valutazione fossero predisposte "al fine di uniformare i criteri di valutazione delle commissioni d'esame" su tutto il territorio nazionale¹². Le "griglie" ministeriali per le prove scritte esordirono nel 2020. Per la prova orale, la prima "griglia" nazionale obbligatoria fu allegata all'ordinanza ministeriale 16 maggio 2020, n. 10, emanata nel contesto dell'emergenza sanitaria da Covid-19, che annullò le prove scritte e concentrò l'esame in un unico colloquio orale. La relazione tecnica allegata all'ordinanza motivava l'introduzione della "griglia" affermando che, in assenza delle prove scritte, le commissioni sarebbero state "del tutto prive di criteri oggettivi nazionali, rischiando in tal modo di snaturare le caratteristiche peculiari dell'esame di Stato conclusivo del secondo ciclo"¹³. L'ordinanza scompose il punteggio del colloquio in indicatori e descrittori di livello definiti a livello centrale, segnando il passaggio da parametri orientativi a criteri vincolanti.

Va considerato che tra il 2020 e il 2026 muta in maniera rilevante la struttura del colloquio. Nel 2020, in seguito alla sospensione delle prove scritte, l'esame viene concentrato in un'unica prova orale, alla quale è affidato un

¹⁰ Legge 10 dicembre 1997, n. 425; Decreto del Presidente della Repubblica 23 luglio 1998, n. 323.

¹¹ Sebbene nei documenti ministeriali tali strumenti siano denominati "griglie di valutazione", la loro struttura li riconduce più propriamente alla categoria delle rubriche valutative, poiché articolano il giudizio in indicatori, livelli, descrittori e punteggi. Nel presente contributo si userà pertanto il termine rubriche in senso docimologico, impiegando "griglie" quando si richiama la denominazione adottata dalle fonti ministeriali.

¹² Decreto legislativo 13 aprile 2017, n. 62, art. 17, commi 5-6.

¹³ Ordinanza ministeriale 16 maggio 2020, n. 10, Allegato B: Griglia di valutazione della prova orale.

peso eccezionalmente elevato e una funzione sostitutiva dell'intero impianto d'esame. In quel contesto la "griglia" ministeriale risponde all'esigenza di rendere comparabile una prova divenuta oramai centrale e totalizzante. Negli anni successivi, con il progressivo ripristino delle prove scritte, il colloquio torna a essere una componente dell'esame, ma conserva la rubrica nazionale come dispositivo di regolazione del giudizio. Dal 2023 al 2025 si stabilizza un modello in cui l'orale mantiene una funzione di raccordo tra discipline, percorso scolastico, esperienze personali e cittadinanza, anche in relazione all'insegnamento dell'educazione civica. Nel 2026, invece, l'assetto cambia in modo più netto: il colloquio viene centrato su quattro discipline e si riduce il riferimento al dialogo ampio tra saperi; scompare, nella rubrica, l'indicatore relativo alla cittadinanza attiva; viene introdotta la valutazione del grado di maturazione personale, autonomia e responsabilità. Un ulteriore cambiamento va collocato nel contesto del dibattito pubblico emerso nel 2025, quando alcune ragazze e alcuni ragazzi già in condizione di superare l'esame grazie ai crediti scolastici e ai risultati conseguiti nelle prove scritte hanno rifiutato deliberatamente di sostenere il colloquio orale come forma di protesta contro il sistema valutativo. Episodi di questo tipo non avevano in precedenza assunto analoga visibilità pubblica nell'esame di Stato e, nel 2025, non hanno impedito ai soggetti coinvolti di conseguire il diploma. Dal 2026, la normativa¹⁴ afferma che l'esame è validamente sostenuto solo se vengono svolte regolarmente tutte le prove. Ne consegue che il rifiuto deliberato di sostenere il colloquio non è più configurabile come prestazione valutabile con punteggio nullo o minimo, ma può incidere sul conseguimento stesso del titolo.

4. Dalla valutazione implicita ai criteri espliciti: il ruolo delle rubriche

La ricostruzione storica mostra come la valutazione del colloquio orale sia stata a lungo affidata a criteri impliciti, variabili e spesso riconducibili a giudizi complessivi sulla maturità culturale e personale del candidato. La progressiva introduzione di dispositivi valutativi articolati in indicatori, livelli e descrittori può essere letta come il tentativo di ridurre la discrezionalità del giudizio e di rendere più trasparenti i criteri di attribuzione del punteggio. In questa prospettiva, è dunque necessario chiarire che cosa sia una rubrica

¹⁴ Decreto legislativo 13 aprile 2017, n. 62, art. 17, comma 2-bis, come modificato dal decreto-legge 9 settembre 2025, n. 127, convertito con modificazioni dalla legge 30 ottobre 2025, n. 164; cfr. anche OM 26 marzo 2026, n. 54.

valutativa, quali funzioni possa svolgere e quali rischi presenti quando viene utilizzata in un contesto sommativo e certificativo come l'esame di Stato.

Sebbene oggi il termine sia ampiamente utilizzato nel campo della valutazione formativa, la parola rubrica rinvia a una storia semantica precedente al suo impiego docimologico. Deriva, infatti, dal latino *ruber*, che rinvia al colore rosso, e rimanda alle intestazioni che nei manoscritti medievali segnalavano, spesso con lettere rosse, le principali sezioni del testo. Questa origine è rilevante anche sul piano valutativo: la rubrica nasce come dispositivo che ordina, evidenzia e rende interpretabili informazioni significative (Ellerani, 2006; Popham, 1997; Robasto, 2022). Trasferita nel campo della valutazione educativa, tale funzione consiste nel rendere espliciti gli aspetti rilevanti di una prestazione, consentendo a docenti e studenti di riconoscere che cosa viene valutato, in base a quali criteri e secondo quali livelli di qualità.

In termini generali, una rubrica può essere definita come uno strumento che articola le aspettative relative a un compito o a una prestazione attraverso criteri di valutazione e descrizioni dei livelli di qualità (Brookhart, 2018; Brookhart e Chen, 2015; Jonsson e Svingby, 2007; Popham, 1997). Popham (1997) ne individua tre elementi essenziali: i criteri valutativi, ossia gli aspetti che contano in una determinata prestazione e che consentono di riconoscere un lavoro appropriato da uno non adeguato; le definizioni di qualità, cioè la descrizione dei diversi gradi con cui tali criteri possono manifestarsi; una strategia di attribuzione del punteggio, che può assumere forma olistica o analitica. Questa impostazione permette di distinguere la rubrica da una semplice griglia di correzione: non si tratta soltanto di attribuire punteggi, ma di descrivere la qualità di una prestazione in rapporto a dimensioni significative (Capperucci, 2016).

In base al grado di articolazione del giudizio, la letteratura distingue tra rubriche olistiche e rubriche analitiche. Le prime restituiscono un giudizio complessivo sulla qualità della prestazione; le seconde, invece, scompongono la prestazione in più dimensioni o criteri, ciascuno dei quali viene descritto e valutato separatamente. Come osservano Jonsson e Svingby (2007), in una rubrica olistica il valutatore formula un giudizio globale sulla performance, mentre in un'analitica assegna un livello o un punteggio a ciascuna dimensione del compito. La rubrica analitica risulta particolarmente utile poiché consente di discriminare tra aspetti diversi della prestazione, individuare punti di forza e aree di miglioramento e ridurre la variabilità del giudizio tra valutatori (Jonsson e Svingby, 2007; Robasto, 2022, 2023).

Un'altra distinzione riguarda il grado di specificità: le rubriche specifiche sono costruite per un determinato compito o una particolare prestazione, quelle generali, invece, descrivono dimensioni trasversali applicabili a più

compiti o prestazioni simili (Capperucci, 2016; Jonsson e Svingby, 2007; Robasto, 2022). Nel primo caso si privilegia l'aderenza al compito, come avviene per una rubrica pensata per la valutazione di una specifica prova; nel secondo, si valorizza la possibilità di osservare e valutare dimensioni ricorrenti di una competenza in contesti diversi, come nel caso di una rubrica progettata per la valutazione di un testo scritto (Castoldi, 2019).

Gli elementi costitutivi di una rubrica, pur denominati in modo non sempre uniforme, possono essere ricondotti ad alcuni nuclei comuni: criteri, livelli, descrittori, indicatori e àncore. I criteri individuano gli aspetti rilevanti da considerare nella valutazione; i descrittori precisano le caratteristiche qualitative dei livelli di prestazione; gli indicatori esplicitano le evidenze osservabili che permettono di riconoscere la presenza o il grado di manifestazione di un criterio; le àncore, infine, consistono in esempi o campioni di prestazione utili a rendere più concreta l'interpretazione dei livelli e a favorire una maggiore coerenza tra valutatori (Brookhart, 2013; Brookhart e Chen, 2015; Ellerani, 2006; Jonsson e Svingby, 2007; Robasto, 2022).

Tra i vantaggi delle rubriche, soprattutto nella valutazione di prestazioni complesse, vi è la possibilità di rendere più coerente il giudizio: le rubriche possono migliorare l'affidabilità della valutazione soprattutto quando sono analitiche, riferite in modo sufficientemente specifico al dominio valutato e accompagnate da esempi di prestazione o da momenti di confronto tra valutatori (Jonsson e Svingby, 2007). Tale funzione dipende però dalla qualità dello strumento: criteri chiari e focalizzati, insieme a descrizioni dei livelli di prestazione, sono condizioni essenziali perché la rubrica produca informazioni valutative significative (Brookhart e Chen, 2015). La rubrica, tuttavia, non garantisce di per sé una valutazione valida: la pertinenza del giudizio dipende dalla coerenza tra ciò che si intende accertare, il compito proposto, i criteri selezionati e i descrittori dei livelli (Jonsson e Svingby, 2007; Robasto, 2022). In altri termini, una rubrica può rendere più stabile e comparabile il giudizio senza renderlo necessariamente più pertinente se gli aspetti osservati non corrispondono effettivamente alla competenza o alla prestazione che si intende accertare.

Un secondo aspetto riguarda la trasparenza: la rubrica rende visibili i criteri di qualità della prestazione e permette allo studente di comprendere meglio che cosa è richiesto, quali dimensioni saranno considerate e come la propria prestazione sarà interpretata. In questa prospettiva, la rubrica non è soltanto uno strumento di valutazione, ma anche un dispositivo di apprendimento: può sostenere il feedback, l'autovalutazione e la regolazione dell'apprendimento, poiché aiuta lo studente a monitorare il proprio lavoro e a riconoscere gli aspetti da migliorare (Andrade e Du, 2005; Brookhart, 2018;

Brookhart e Chen, 2015; Andrade e Brookhart, 2026; Panadero, Alonso-Tapi e Reche, 2013; Panadero e Jonsson, 2013).

In ogni caso, è opportuno distinguere tra uso formativo e uso sommativo-certificativo della rubrica. Nell'uso formativo, la rubrica è impiegata durante il percorso di apprendimento per chiarire i criteri di qualità, offrire feedback, sostenere l'autovalutazione e favorire il miglioramento progressivo della prestazione; non serve soltanto a registrare un esito, ma accompagna lo studente nella comprensione dei possibili sviluppi del proprio lavoro (Panadero, Alonso-Tapi e Reche, 2013; Panadero e Jonsson, 2013). Nell'uso sommativo o certificativo, invece, la rubrica sostiene l'attribuzione di un giudizio finale o di un punteggio. Questo secondo impiego è significativo nei contesti valutativi a forte rilevanza certificativa, come gli esami conclusivi, nei quali assumono centralità la comparabilità dei giudizi, l'affidabilità dell'attribuzione del punteggio e la possibilità di giustificarlo sulla base di evidenze osservabili (Jonsson e Svingby, 2007; Brookhart e Chen, 2015; Andrade e Brookhart, 2026). Le due funzioni non sono necessariamente alternative: anche una rubrica impiegata a fini certificativi può mantenere una funzione descrittiva e orientativa se chiarisce il significato del giudizio attribuito in termini di qualità della prestazione. Al contrario, quando si limita a stabilire corrispondenze tra livelli, aggettivi e punteggi, rischia di ridursi a un dispositivo di classificazione, più vicino a una tabella di conversione che a uno strumento di interpretazione della prestazione. Questa criticità richiama la distinzione tra rubriche descrittive, fondate su criteri e descrizioni dei livelli di qualità e semplici schemi di punteggio o scale valutative (Brookhart e Chen, 2015). In entrambi gli usi, formativo e certificativo, il valore della rubrica dipende dalla qualità della sua costruzione. I descrittori dovrebbero quindi chiarire che cosa caratterizza qualitativamente un livello rispetto a un altro, evitando formulazioni generiche o meramente quantitative. In altri termini, una progressione affidata solo ad avverbi come "parzialmente", "abbastanza", "adeguatamente" o "pienamente" rischia infatti di lasciare implicito ciò che dovrebbe invece essere descritto: che cosa lo studente sa fare, con quale grado di autonomia, pertinenza, complessità e padronanza. In questa direzione, Tierney e Simon (2004) richiamano la necessità di mantenere coerenza nel linguaggio dei descrittori lungo i diversi livelli della scala, affinché la rubrica descriva effettivamente la qualità della prestazione.

5. La valutazione del colloquio orale

Nel caso del colloquio orale, la questione dei criteri valutativi assume un rilievo specifico, poiché la prova non coincide con una semplice verifica di conoscenze, ma si propone piuttosto di valutare una prestazione complessa, situata e interattiva. La letteratura sull'*oral assessment* evidenzia che una prova orale è una forma valutativa multidimensionale, che può variare per tipo di contenuto accertato, grado di interazione tra esaminatore e candidato, autenticità e struttura del compito, ruolo dell'esaminatore e peso attribuito alla dimensione propriamente orale della prestazione (Joughin, 1998). Gli studi sulle rubriche per prestazioni orali, pur riferendosi a contesti non sovrapponibili al colloquio d'esame, confermano l'utilità di criteri analitici per descrivere dimensioni osservabili della prestazione, come contenuto, struttura, interazione e modalità comunicative (van Ginckel et al., 2017).

Queste indicazioni consentono di leggere le “griglie” ministeriali per il colloquio dell'esame di Stato non solo come strumenti amministrativi di attribuzione del punteggio, ma come dispositivi che tentano di rendere analiticamente osservabile una prestazione orale complessa. Presentano infatti alcuni tratti propri delle rubriche valutative, nella misura in cui articolano la valutazione in indicatori, livelli, descrittori e punteggi. Tale lettura si pone in continuità con i cambiamenti descritti nei paragrafi precedenti: il modello del 2020 ha introdotto per la prova orale una “griglia” nazionale a punteggio fisso, con indicatori e descrittori di livello definiti, al fine di ridurre i margini di discrezionalità locale che avevano caratterizzato la valutazione del colloquio.

La presenza di un dispositivo, tuttavia, non elimina automaticamente l'inaffidabilità del giudizio. In particolare, è la qualità dei descrittori a determinare la capacità dello strumento di orientare in maniera affidabile la valutazione. Descrittori troppo generici o differenziati attraverso aggettivi come “parziale”, “adeguato” o “pieno” rischiano di non rendere visibile il passaggio qualitativo da un livello all'altro. Una rubrica efficace dovrebbe descrivere quali aspetti della prestazione consentono di riconoscere un diverso grado di padronanza: la pertinenza della risposta, la capacità di collegare conoscenze, la qualità dell'argomentazione, l'uso del lessico disciplinare e il grado di autonomia nella rielaborazione.

In questa prospettiva, il rischio non è l'uso della “griglia” in sé, ma la sua riduzione a meccanismo di attribuzione del punteggio. Se i descrittori sono generici, se la progressione tra livelli è affidata a etichette poco informative, se il rapporto tra indicatori e prestazione osservata resta debole, il dispositi-

vo rischia di risolversi in uno strumento impiegato allo scopo di giustificare una valutazione preconfezionata. Il rischio è quello del “voto mascherato”: un dispositivo apparentemente descrittivo che finisce per tradurre i punteggi assegnati in etichette qualitative, senza rendere davvero trasparente il processo interpretativo che conduce al giudizio. Tale rischio richiama la critica di Popham (1997) alle rubriche formalmente strutturate ma poco utili per comprendere la qualità della prestazione, e la distinzione proposta da Brookhart e Chen (2015) tra rubriche descrittive, fondate su criteri e livelli di qualità, e semplici schemi di punteggio.

6. Analisi delle rubriche ministeriali per il colloquio orale dal 2020 al 2026

I cambiamenti nelle rubriche ministeriali per il colloquio orale dal 2020 (fig. 1) al 2026 (fig. 2) mostrano i passaggi di un dispositivo nato nel contesto emergenziale della pandemia e successivamente incorporato nella struttura ordinaria dell'esame. Gli indicatori riguardano l'acquisizione di contenuti e metodi disciplinari, la capacità di utilizzare e collegare conoscenze, l'argomentazione critica e personale, la padronanza lessicale e semantica e la capacità di analisi e comprensione della realtà in chiave di cittadinanza attiva. Dal 2020 al 2025 le rubriche presentano una forte continuità: mantengono gli stessi indicatori e descrittori, varia la distribuzione dei punteggi.

La rubrica del 2026 introduce la discontinuità più marcata: gli indicatori passano da cinque a quattro, il riferimento alle discipline viene ristretto alle quattro discipline oggetto del colloquio, la padronanza lessicale e semantica viene accorpata all'uso e raccordo delle conoscenze, scompare l'indicatore sulla cittadinanza attiva e viene introdotto un indicatore sul “grado di maturazione personale, di autonomia e di responsabilità”. Va considerato che dal 2026 il colloquio è meno orientato al dialogo tra i diversi saperi disciplinari e maggiormente centrato sulle discipline estratte.

Nel complesso, le rubriche evidenziano due tendenze. Da un lato, vi è il tentativo di ridurre la discrezionalità del giudizio mediante uno strumento comune, vincolante e articolato in indicatori, livelli, descrittori e punteggi. Dall'altro, permane la difficoltà metodologica nel proporre rubriche autenticamente descrittive, capaci di rendere osservabili le dimensioni della prestazione e sostenere il confronto intersoggettivo nella commissione. Come evidenziato, la letteratura scientifica sulle rubriche evidenzia l'importanza di criteri pertinenti, di dimensioni non sovrapponibili, di livelli qualitativamente differenti di prestazione, dell'esplicitazione delle evidenze osservabili per

agevolare, oltre all'attribuzione del punteggio, la comprensione del giudizio. La presenza di una rubrica non garantisce dunque, di per sé, validità e affidabilità: tali qualità dipendono in primo luogo dalla chiarezza di indicatori e descrittori.

Un primo limite ricorrente riguarda la formulazione degli indicatori. Nelle rubriche 2020-2025, alcuni indicatori distinguono dimensioni diverse della prestazione: contenuti e metodi, collegamenti, argomentazione, linguaggio, cittadinanza. Tuttavia, anche in questa struttura più analitica, le dimensioni restano ampie e scarsamente operazionalizzate. La rubrica 2026, riducendo il numero degli indicatori e accorpendo dimensioni precedentemente separate, aggrava il problema. Il secondo indicatore, per esempio, somma la capacità di utilizzare e raccordare conoscenze, la padronanza lessicale e semantica, il riferimento al linguaggio tecnico o settoriale e, eventualmente, alla lingua straniera. Si tratta di aspetti rilevanti, ma distinti: la capacità di costruire raccordi concettuali non coincide né con la padronanza del lessico disciplinare né con la qualità semantica dell'esposizione. Ancora più problematica è la formulazione del quarto indicatore del 2026, che mette assieme maturazione personale, autonomia e responsabilità. L'ampiezza di tali categorie rende difficile identificarne manifestazioni osservabili nel tempo limitato e nel setting fortemente regolato del colloquio.

Un secondo limite, connesso al primo, riguarda i descrittori. Nelle rubriche in uso fino al 2025, per l'indicatore relativo all'acquisizione dei contenuti e dei metodi, la progressione tra livelli è basata su formule come "in modo parziale e incompleto", "in modo corretto e appropriato", "in maniera completa" e "in maniera completa e approfondita". Analogamente, per la capacità di utilizzare e collegare le conoscenze, il passaggio tra livelli è affidato a espressioni quali "con difficoltà e in modo stentato", "adeguati collegamenti", "trattazione pluridisciplinare articolata" e "trattazione pluridisciplinare ampia e approfondita". In questi casi, la rubrica indica una scala graduata senza chiarire quali evidenze osservabili consentano di distinguere, nel corso del colloquio, un collegamento "adeguato" da uno "articolato", o una trattazione "articolata" da una "ampia e approfondita". Lo stesso limite si aggrava nel 2026. La scelta di accorpare dimensioni eterogenee nello stesso indicatore e l'introduzione del "grado di maturazione, di autonomia e di responsabilità" non incide positivamente sull'affidabilità dei descrittori. A titolo esemplificativo, si segnala come la scala ordinale per la valutazione dell'ultimo indicatore rappresenti un caso particolarmente evidente di voto mascherato da descrizione: inizia con "molto parziale", procede con "limitato", "apprezzabile", "alto" e si conclude con "elevato".

<i>Indicatori</i>	<i>Livelli</i>	<i>Descrittori</i>	<i>Punti</i>
Acquisizione dei contenuti e dei metodi delle diverse discipline del curriculum, con particolare riferimento a quelle d'indirizzo	I	Non ha acquisito i contenuti e i metodi delle diverse discipline, o li ha acquisiti in modo estremamente frammentario e lacunoso	1-2
	II	Ha acquisito i contenuti e i metodi delle diverse discipline in modo parziale e incompleto, utilizzandoli in modo non sempre appropriato	3-5
	III	Ha acquisito i contenuti e utilizza i metodi delle diverse discipline in modo corretto e appropriato	6-7
	IV	Ha acquisito i contenuti delle diverse discipline in maniera completa e utilizza in modo consapevole i loro metodi	8-9
	V	Ha acquisito i contenuti delle diverse discipline in maniera completa e approfondita e utilizza con piena padronanza i loro metodi	10
Capacità di utilizzare le conoscenze acquisite e di collegarle tra loro	I	Non è in grado di utilizzare e collegare le conoscenze acquisite o lo fa in modo del tutto inadeguato	1-2
	II	È in grado di utilizzare e collegare le conoscenze acquisite con difficoltà e in modo stentato	3-5
	III	È in grado di utilizzare correttamente le conoscenze acquisite, istituendo adeguati collegamenti tra le discipline	6-7
	IV	È in grado di utilizzare le conoscenze acquisite collegandole in una trattazione pluridisciplinare articolata	8-9
	V	È in grado di utilizzare le conoscenze acquisite collegandole in una trattazione pluridisciplinare ampia e approfondita	10
Capacità di argomentare in maniera critica e personale, elaborando i contenuti acquisiti	I	Non è in grado di argomentare in maniera critica e personale, o argomenta in modo superficiale e disorganico	1-2
	II	È in grado di formulare argomentazioni critiche e personali solo a tratti e solo in relazione a specifici argomenti	3-5
	III	È in grado di formulare semplici argomentazioni critiche e personali, con una corretta rielaborazione dei contenuti acquisiti	6-7
	IV	È in grado di formulare articolate argomentazioni critiche e personali, rielaborando efficacemente i contenuti acquisiti	8-9
	V	È in grado di formulare ampie e articolate argomentazioni critiche e personali, rielaborando con originalità i contenuti acquisiti	10
Ricchezza e padronanza lessicale e semantica, con specifico riferimento al linguaggio tecnico e/o di settore, anche in lingua straniera	I	Si esprime in modo scorretto o stentato, utilizzando un lessico inadeguato	1
	II	Si esprime in modo non sempre corretto, utilizzando un lessico, anche di settore, parzialmente adeguato	2
	III	Si esprime in modo corretto utilizzando un lessico adeguato, anche in riferimento al linguaggio tecnico e/o di settore	3
	IV	Si esprime in modo preciso e accurato utilizzando un lessico, anche tecnico e settoriale, vario e articolato	4
	V	Si esprime con ricchezza e piena padronanza lessicale e semantica, anche in riferimento al linguaggio tecnico e/o di settore	5
Capacità di analisi e comprensione della realtà in chiave di cittadinanza attiva a partire dalla riflessione sulle esperienze personali	I	Non è in grado di analizzare e comprendere la realtà a partire dalla riflessione sulle proprie esperienze, o lo fa in modo inadeguato	1
	II	È in grado di analizzare e comprendere la realtà a partire dalla riflessione sulle proprie esperienze con difficoltà e solo se guidato.	2
	III	È in grado di compiere un'analisi adeguata della realtà sulla base di una corretta riflessione sulle proprie esperienze personali.	3
	IV	È in grado di compiere un'analisi precisa della realtà sulla base di una attenta riflessione sulle proprie esperienze personali.	4
	V	È in grado di compiere un'analisi approfondita della realtà sulla base di una riflessione critica e consapevole sulle proprie esperienze personali	5

Figura 1 – Rubrica 2020

<i>Indicatori</i>	<i>Livelli</i>	<i>Descrittori</i>	<i>Punti</i>
Acquisizione dei contenuti e dei metodi delle quattro discipline oggetto del colloquio	I	Non ha acquisito i contenuti e i metodi delle diverse discipline, o li ha acquisiti in modo estremamente frammentario e lacunoso	0,5-1,0
	II	Ha acquisito i contenuti e i metodi delle diverse discipline in modo parziale e/o incompleto, e li utilizza in modo non sempre appropriato	1,5-2,5
	III	Ha acquisito i contenuti e utilizza i metodi delle diverse discipline in modo corretto e appropriato	3-3,5
	IV	Ha acquisito i contenuti delle diverse discipline in maniera completa e utilizza in modo consapevole i relativi metodi	4-4,5
	V	Ha acquisito i contenuti delle diverse discipline in maniera completa e approfondita e utilizza con piena padronanza i relativi metodi	5,0
Capacità di utilizzare e ricordare le conoscenze acquisite; padronanza lessicale e semantica, anche con riferimento al linguaggio tecnico e/o di settore (eventualmente anche in lingua straniera)	I	Non è in grado di utilizzare e ricordare le conoscenze acquisite o lo fa in modo del tutto inadeguato. Si esprime in modo scorretto	0,5-1,0
	II	È in grado di utilizzare e ricordare le conoscenze acquisite con difficoltà e solo se guidato. Si esprime in modo non sempre corretto, utilizzando un lessico, anche di settore, parzialmente adeguato	1,5-2,5
	III	È in grado di utilizzare correttamente le conoscenze acquisite, istituendo adeguati raccordi tra le discipline. Si esprime utilizzando un lessico complessivamente corretto, anche in riferimento al linguaggio tecnico e/o di settore	3,0-3,5
	IV	È in grado di utilizzare le conoscenze acquisite raccordandole in una trattazione pluridisciplinare articolata. Si esprime in modo preciso e accurato utilizzando un lessico, anche tecnico e settoriale, vario e preciso	4,0-4,5
	V	È in grado di utilizzare le conoscenze acquisite raccordandole in una trattazione pluridisciplinare ampia e approfondita. Si esprime con ricchezza e piena padronanza lessicale e semantica, anche in riferimento al linguaggio tecnico e/o di settore	5,0
Capacità di argomentare in modo critico e personale	I	Non è in grado di argomentare in maniera critica e personale, o argomenta in modo superficiale e disorganico	0,5-1,0
	II	È in grado di formulare argomentazioni critiche e personali solo a tratti e/o solo in relazione a specifici argomenti	1,5-2,5
	III	È in grado di formulare semplici argomentazioni critiche e personali, rielaborando correttamente i contenuti acquisiti	3,0-3,5
	IV	È in grado di formulare articolate argomentazioni critiche e personali, rielaborando efficacemente i contenuti acquisiti	4,0-4,5
	V	È in grado di formulare ampie e articolate argomentazioni critiche e personali, rielaborando con originalità i contenuti acquisiti	5,0
Grado di maturazione personale, di autonomia e di responsabilità raggiunto al termine del percorso di studio	I	Ha raggiunto un grado di maturazione molto parziale e un livello di autonomia e responsabilità incompleto	0,5-1,0
	II	Ha raggiunto un limitato grado di maturazione e di autonomia; necessità di guida e di supporto per gestire scelte e responsabilità	1,5-2,5
	III	Ha raggiunto un apprezzabile livello di maturazione; è in grado di assumere decisioni autonome e gestire con sicurezza scelte personali	3,0-3,5
	IV	Ha raggiunto un alto grado di maturazione, autonomia e responsabilità; è capace di riflettere criticamente sulle proprie scelte e sul proprio agire	4,0-4,5
	V	Ha raggiunto un elevato grado di autonomia e maturazione personale; sa gestire responsabilità significative in modo esemplare per gli altri	5,0

Figura 2 – Rubrica 2026

Un terzo limite riguarda la debolezza delle evidenze osservabili. Una prestazione orale complessa richiede criteri capaci di descrivere comportamenti valutabili nel corso dell'interazione: pertinenza delle risposte, organizzazione del discorso, uso delle fonti e dei contenuti, costruzione di nessi, capacità di sostenere un'argomentazione, gestione delle sollecitazioni, consapevolezza metacognitiva. Le rubriche ministeriali richiamano alcune di queste dimensioni, ma spesso a un livello di generalità tale da non orientare in modo sufficientemente preciso la raccolta delle evidenze.

Un quarto limite riguarda il rapporto tra funzione certificativa e funzione descrittiva. In un esame conclusivo è legittimo che la rubrica sostenga l'attribuzione di un punteggio. Tuttavia, anche in un uso sommativo e certificativo, la rubrica deve conservare una funzione descrittiva. Le rubriche esaminate mostrano invece una tendenza a privilegiare la distribuzione amministrativa del punteggio rispetto alla descrizione analitica della prestazione. La struttura a cinque livelli produce un'apparenza di precisione, ma non aumenta la qualità informativa del giudizio.

La svolta del 2026 rende tali limiti ancora più evidenti. L'introduzione della maturazione personale, dell'autonomia e della responsabilità riattiva categorie storicamente presenti nell'esame di maturità, ma le colloca dentro un dispositivo standardizzato che non chiarisce le condizioni della loro osservabilità. Infine, l'imposizione delle discipline da trattare e la clausola relativa al regolare svolgimento di tutte le prove rafforzano il carattere vincolante e performativo del colloquio. Il soggetto valutato ha l'obbligo di mostrarsi maturo, autonomo e responsabile in un contesto in cui lo spazio di autodeterminazione è fortemente limitato dalla struttura dell'esame, da imposizioni contenutistiche, dalle domande della commissione, dai tempi e dalle condizioni istituzionali della prova. Autonomia e responsabilità vengono valutate da un dispositivo che le limita fortemente.

Conclusioni

In conclusione, le rubriche ministeriali per il colloquio orale dal 2020 al 2026 rappresentano un passaggio importante nella storia della valutazione dell'esame di Stato, perché rendono esplicito il tentativo di governare la discrezionalità del giudizio attraverso criteri comuni. Tuttavia, evidenziano ampi margini di miglioramento: gli indicatori sono scarsamente indicativi e i descrittori non forniscono descrizioni coerenti. Va nondimeno considerato che i problemi non dipendono soltanto da limiti tecnici nella costruzione dello strumento, ma rinviano all'ambiguità della richiesta valutativa alla base

del colloquio orale: valutare, nello stesso momento, conoscenze disciplinari, raccordi tra saperi, capacità argomentativa, padronanza lessicale, cittadinanza, autonomia, responsabilità e maturazione. Una rubrica può ordinare tale complessità, ma non può risolvere da sola la genericità e l'indeterminatezza dell'oggetto valutato. Senza una definizione chiara di cosa il colloquio intenda effettivamente accertare, il rischio è che la rubrica non renda più fondato il giudizio, ma nasconda sotto una cappa di oggettività – insostenibile sul piano docimologico – l'arbitrarietà di una valutazione finalizzata più a classificare gli individui che a rendere conto dei loro apprendimenti.

Riferimenti bibliografici

- Andrade, H. L., Brookhart, S. M. (2026), *Using Rubrics for Teaching and Learning*. New York: Routledge, <https://doi.org/10.4324/9781003582649>.
- Andrade, H. L., Du, Y. (2005), "Student Perspectives on Rubric-Referenced Assessment", *Educational & Counseling Psychology Faculty Scholarship*, 10 (3), pp. 1-9.
- Brookhart, S. M. (2013), *How to Create and Use Rubrics for Formative Assessment and Grading*. Alexandria: ASCD.
- Brookhart, S. M. (2018), "Appropriate Criteria: Key to Effective Rubrics", *Frontiers in Education*, 3, 22, <https://doi.org/10.3389/educ.2018.00022>.
- Brookhart, S. M., Chen, F. (2015), "The quality and effectiveness of descriptive rubrics", *Educational Review*, 67 (3), pp. 343-368, <https://www.tandfonline.com/doi/full/10.1080/00131911.2014.929565>.
- Capperucci, D. (2016), "L'uso delle rubriche valutative per la certificazione delle competenze. Il modello Va.R.C.Co.", *Form@re – Open Journal per la formazione in rete*, 16 (1), pp. 133-151, <https://doi.org/10.13128/formare-17936>.
- Castoldi, M. (2019), *Rubriche valutative. Guidare l'espressione del giudizio*. Torino: Utet.
- Cives, G. (a cura di) (1994), *La scuola italiana dall'Unità ai nostri giorni*. Firenze: La Nuova Italia, 1ª ed., 4ª rist.
- Corsini, C. (2023), *La valutazione che educa. Liberare insegnamento e apprendimento dalla tirannia del voto*. Milano: FrancoAngeli.
- De Landsheere, G. (1971), *Evaluation continue et examens. Précis de docimologie*. Bruxelles: Labor.
- Ellerani, P. G. (2006), "Per una valutazione 'autentica'", *Innovazione educativa*, 2, pp. 50-56.
- Galfré, M. (2020), Il '68 e il rinnovamento dell'insegnamento nella scuola media superiore. In R. Sani, A. Ascenzi (a cura di), *L'innovazione pedagogica e didattica nella scuola italiana del secondo Novecento*. Roma: Studium, pp. 510-523.
- Jonsson, A., Svingby, G. (2007), "The use of scoring rubrics: Reliability, validity and educational consequences", *Educational Research Review*, 2 (2), pp. 130-144, <https://doi.org/10.1016/j.edurev.2007.05.002>.

- Joughin, G. (1998), “Dimensions of Oral Assessment”, *Assessment & Evaluation in Higher Education*, 23 (4), pp. 367-378, <https://doi.org/10.1080/0260293980230404>.
- Morandi, M. (2019), “Il Sessantotto della docimologia”, *Educazione. Giornale di pedagogia critica*, VIII, 2, pp. 33-50.
- Morandi, M. (2025), *Storia critica del voto scolastico*. Brescia: Scholè.
- Panadero, E., Alonso-Tapia, J., Reche, E. (2013), “Rubrics vs. Self-assessment scripts effect on self-regulation, performance and self-efficacy in pre-service teachers”, *Studies in Educational Evaluation*, 39 (3), pp. 125-132, <https://doi.org/10.1016/j.stueduc.2013.04.001>.
- Panadero, E., Jonsson, A. (2013), “The use of scoring rubrics for formative assessment purposes revisited: A review”, *Educational Research Review*, 9, pp. 129-144.
- Piéron, H. (1963), *Examens et docimologie*. Paris: Presses Universitaires de France.
- Popham, W. J. (1997), “What’s wrong – And what’s right – With Rubrics”, *Educational Leadership*, 55 (2), pp. 72-75.
- Robasto, D. (2022), Le rubriche di valutazione nell’Higher Education. In P. Ricchiardi, E. Torre, R. Trincherò (a cura di), *Percorsi di ricerca didattica e docimologia. Studi in onore di Cristina Coggi*. Milano: FrancoAngeli, pp. 107-122.
- Robasto, D. (2023), “Quanto mi dai? O Come miglioro? Costruire rubriche valutative per supportare la valutazione formativa”, *Lifelong Lifewide Learning*, 26, pp. 87-98.
- Tierney, R., Simon, R. (2004), “What’s still wrong with rubrics: Focusing on the consistency of performance criteria across scale levels”, *Practical Assessment, Research & Evaluation*, 9 (2), pp. 1-10.
- Van Ginkel, S., Laurentzen, R., Mulder, M., Mononen, A., Kytä, J., Kortelainen, M. J. (2017), “Assessing oral presentation performance: Designing a rubric and testing its validity with an expert group”, *Journal of Applied Research in Higher Education*, 9 (3), pp. 474-486, <https://doi.org/10.1108/JARHE-02-2016-0012>.