

Beyond formative assessment: Construction and validation of the Teachers' Assessment Strategies Scale (StraVI)

Oltre la valutazione formativa: Costruzione e validazione della scala delle Strategie Valutative degli Insegnanti (StraVI)

Irene Dora Maria Scierri

Dipartimento di Formazione, Lingue, Intercultura, Letterature e Psicologia, Università di Firenze – irene.scierri@unifi.it
<https://orcid.org/0000-0003-4338-0130>

OPEN ACCESS



DOUBLE BLIND PEER REVIEW

ABSTRACT

The study presents the process of construction and validation of the Teachers' Assessment Strategies Scale (StraVI), designed to identify the formative assessment strategies employed by primary and secondary school teachers for in-class assessment. The validation sample consists of 1,545 serving teachers, distributed nationwide. The StraVI scale, subjected to both exploratory and confirmatory factor analyses, demonstrates robust psychometric properties and is delineated into the following dimensions: *Assessment strategies oriented towards improving learning* (S - AfL) and *Assessment strategies oriented towards self-regulation and sustainability of learning* (S - AaL). This instrument addresses a gap in the existing toolkit within the field, by honing in on the specifics of various formative assessment strategies, distinguishing between strategies associated with the *assessment for learning* approach and those linked to the *assessment as learning* approach, while also paying particular attention to *sustainable assessment*.

Lo studio presenta il processo di costruzione e validazione della scala delle Strategie Valutative degli Insegnanti (StraVI), progettata per rilevare le strategie di valutazione di tipo formativo utilizzate dagli insegnanti di scuola primaria e secondaria per la valutazione in classe. Il campione di validazione è costituito da 1.545 docenti in servizio, distribuiti su tutto il territorio nazionale. La scala StraVI, sottoposta ad analisi fattoriale esplorativa e confermativa, mostra buone proprietà psicometriche e si articola nelle seguenti dimensioni: *Strategie valutative orientate al miglioramento degli apprendimenti* (S - AfL) e *Strategie valutative orientate all'autoregolazione e alla sostenibilità dell'apprendimento* (S - AaL). Questo strumento colma un vuoto nel panorama degli strumenti disponibili nel settore, poiché mette a fuoco le specificità delle diverse strategie di valutazione formativa, distinguendo tra strategie legate all'approccio di *assessment for learning* e strategie correlate all'approccio di *assessment as learning*, con una particolare attenzione anche al *sustainable assessment*.

KEYWORDS

Assessment for learning, Assessment as learning, Sustainable assessment, Assessment strategies, Validation Valutazione per l'apprendimento, Valutazione come apprendimento, Valutazione sostenibile, Strategie valutative, Validazione

Citation: Scierri, I.D.M. (2024). Beyond formative assessment: Construction and validation of the Teachers' Assessment Strategies Scale (StraVI). *Formazione & insegnamento*, 22(1), 97-108. https://doi.org/10.7346/-fei-XXII-01-24_11

Copyright: © 2024 Author(s).

License: Attribution 4.0 International (CC BY 4.0).

Conflicts of interest: The Author(s) declare(s) no conflicts of interest.

DOI: https://doi.org/10.7346/-fei-XXII-01-24_11

Submitted: March 16, 2024 • **Accepted:** April 24, 2024 • **Published:** April 30, 2024

Pensa MultiMedia: ISSN 2279-7505 (online)

1. Introduzione

I termini “valutazione sommativa” e “valutazione formativa” sono comunemente impiegati per distinguere le due principali funzioni della valutazione. Nella valutazione sommativa, l’obiettivo principale è quello di classificare o comunque accertare il rendimento degli studenti. Diversamente, la valutazione formativa mira a utilizzare le evidenze derivanti dai risultati degli studenti per informare azioni mirate a facilitare e migliorare il processo di apprendimento (Wiliam & Black, 1996). Numerose ricerche hanno mostrato gli effetti positivi delle pratiche di valutazione formativa su diversi aspetti dell’apprendimento, come i risultati di apprendimento (Batini & Guerra, 2020; Hattie, 2009; Schneider & Preckel, 2017; Wisniewski et al., 2020), il senso di autoefficacia (Panadero et al., 2012; Scierri et al., 2023), la motivazione intrinseca (Meusen-Beekman et al., 2016) e la capacità di autoregolazione dell’apprendimento (Scierri, 2021). Tuttavia, il concetto di valutazione formativa può essere ambiguo e non sempre sono chiare le strategie che vengono utilizzate nell’applicazione pratica di questo approccio, il che può portare a interpretazioni non uniformi (Schellekens et al., 2021; Vertecchi, 2023). È quindi fondamentale giungere a una definizione chiara delle valutazioni di tipo formativo, per poter documentare l’efficacia di differenti approcci, implementarli correttamente e sintetizzare i risultati delle ricerche (Bennett, 2011). In questa prospettiva, risulta utile andare oltre la valutazione formativa per focalizzarsi su alcuni approcci specifici come l’*assessment for learning* (AfL), l’*assessment as learning* (AaL) e il *sustainable assessment*. Tuttavia, un’analisi della letteratura esistente non ha rilevato strumenti in grado di cogliere appieno le specificità, soprattutto operative, di questi approcci valutativi. In particolare, l’AaL e il *sustainable assessment*, che assumono un ruolo cruciale nel promuovere le capacità metacognitive e di autoregolazione dell’apprendimento, non sono stati adeguatamente indagati. Questa mancanza rappresenta un ostacolo per la conduzione di ricerche empiriche atte a valutare la diffusione e gli effetti di questi approcci. Per colmare questa lacuna, sono stati sviluppati due strumenti¹ che si concentrano sulle concezioni e sulle strategie valutative degli insegnanti, tenendo conto anche delle concezioni e delle pratiche incentrate sullo sviluppo dell’autoregolazione e della *sostenibilità dell’apprendimento*².

Il lavoro fa parte di un progetto di ricerca più ampio, condotto utilizzando un approccio *mixed methods*, che esplora le concezioni, le strategie valutative e le percezioni di autoefficacia dei docenti riguardo all’implementazione di strategie di autoregolazione in classe. I principali risultati quantitativi di tale studio, che ha coinvolto un campione più ampio

rispetto agli studi di validazione, sono illustrati in un precedente lavoro (Scierri, 2023). Il presente contributo illustra il processo di costruzione e validazione della scala delle Strategie Valutative degli Insegnanti (StraVI), progettata per rilevare le strategie proprie degli approcci di *assessment for learning* (AfL) e di *assessment as learning* (AaL), quest’ultimo integrato dalla prospettiva del *sustainable assessment*. Il target di riferimento sono gli insegnanti di scuola primaria e secondaria.

2. Le strategie di valutazione formativa: approcci teorici di riferimento

La cornice teorica che ha orientato la costruzione degli item della scala StraVI si fonda sugli approcci valutativi di AfL e AaL, quest’ultimo strettamente connesso al concetto di *sustainable assessment*. Pur rientrando entrambi nella prospettiva formativa o educativa (Corsini, 2023; Galliani, 2009) della valutazione, essi presentano peculiarità distintive. Tra queste, vi sono elementi chiave come la finalità principale per cui la valutazione viene effettuata; il livello di coinvolgimento, partecipazione e autonomia degli studenti nel processo valutativo; e i processi cognitivi e metacognitivi che vengono stimolati. Partendo da questa prospettiva e rimandando ad altri lavori per un maggior approfondimento (Scierri, 2022, 2023), è possibile delineare brevemente le caratteristiche di tali approcci e le strategie valutative che li contraddistinguono.

L’approccio di AfL è comunemente concepito come un processo incentrato sulla ricerca e sull’interpretazione delle prove che studenti e insegnanti impiegano per determinare il progresso degli studenti nel loro percorso di apprendimento, dove quest’ultimi devono arrivare e quali strategie adottare per arrivarci (ARG, 2002). In una definizione più specifica, l’AfL si concentra su un obiettivo di apprendimento specifico o su un insieme di obiettivi strettamente connessi. In questa prospettiva, la valutazione per l’apprendimento comprende tre elementi chiave: la definizione di un obiettivo chiaro; l’identificazione delle lacune tra le competenze attuali degli studenti e l’obiettivo prefissato; e l’individuazione delle strategie e dei passaggi necessari per colmare tali lacune (Crooks, 2011). In breve, la valutazione per l’apprendimento mira a mettere in luce i punti di forza e di debolezza di ogni studente e a fornire a ciascuno un feedback che favorisca l’apprendimento.

Nell’approccio di AaL (Earl, 2013; Trincherò, 2017; Yan & Boud, 2022), il focus si sposta sul contributo dei processi valutativi nel promuovere lo sviluppo delle capacità e dell’impegno degli studenti come soggetti attivi nel processo di apprendimento. Essere protagonisti del proprio percorso di apprendimento richiede la competenza di monitorare il proprio progresso, modificare le strategie di apprendimento secondo le esigenze e assumere una maggiore responsabilità nel processo educativo. Queste capacità sono strettamente connesse con il potenziamento dell’autoregolazione e della metacognizione (Crooks, 2011).

Alcuni autori considerano l’AaL come un sottinsieme dell’AfL (Clark, 2012; Earl, 2013; Lam, 2019), men-

1 Il secondo strumento, denominato Scala delle Concezioni Valutative degli Insegnanti (CoVI), è attualmente in fase di pubblicazione.

2 Il concetto di *sustainable learning* si riferisce a un apprendimento “che duri nel tempo” (Graham et al., 2015). Un apprendimento sostenibile deve focalizzarsi sulle strategie e le abilità che permettono agli studenti di rinnovare, ricostruire, riutilizzare il proprio apprendimento in diverse circostanze, durante le transizioni della vita e in diversi ambiti (Ben-Eliyahu, 2021).

tre altri lo vedono come il culmine di un continuum nella pratica della valutazione per migliorare l'apprendimento (Tomlinson, 2007). In questo lavoro viene adottata la seconda prospettiva, riconoscendo l'AaL come approccio distinto, sebbene strettamente correlato all'AfL. In particolare, l'AaL si distingue per la sua enfasi nel fornire agli studenti opportunità di apprendimento attraverso il coinvolgimento attivo nella valutazione (Yan & Boud, 2022). Ciò che caratterizza l'AaL è quindi la possibilità per gli studenti di acquisire nuove conoscenze e sviluppare le capacità di apprendimento durante l'esecuzione delle attività valutative, una caratteristica che generalmente non si riscontra nell'AfL (Yan & Boud, 2022).

All'interno dell'approccio di AaL è possibile includere anche il *sustainable assessment* (Boud, 2000; Boud & Soler, 2016). Questo approccio pone l'accento, più che sulla finalità della valutazione, su un suo requisito imprescindibile: la *sostenibilità*, ovvero la capacità di preparare gli studenti a rispondere alle loro future esigenze di apprendimento. La valutazione sostenibile si concentra sul concetto di "giudizio valutativo", ossia sull'abilità degli studenti di fornire giudizi sulla qualità del proprio e altrui lavoro. Si tratta di una competenza cruciale, legata alla consapevolezza di ciò che si è capaci di fare, indispensabile sia per avanzare nell'apprendimento – come requisito essenziale per delineare i prossimi passi nel proprio percorso di apprendimento – sia per operare nella società con una buona percezione di fiducia in sé (Boud, 2022). Oltre alla consapevolezza relativa al proprio lavoro, gli studenti devono essere in grado di giudicare il lavoro degli altri, una competenza importante per poter fare affidamento sulle opinioni e sulle azioni altrui, garantendo così la capacità di agire in modo appropriato in qualsiasi situazione (Boud, 2022).

Nella prospettiva delineata, l'AaL implica che le attività valutative contribuiscano alla formazione della capacità di giudizio valutativo degli studenti. Questo approccio è anche in linea con la prospettiva dell'apprendimento lungo tutto l'arco della vita e dell'apprendimento autoregolato.

Le cinque strategie chiave di valutazione formativa offrono un solido fondamento per le pratiche valutative che i docenti possono implementare in classe:

1. chiarire, condividere e comprendere gli obiettivi di apprendimento e i criteri per il conseguimento del risultato (anche attraverso l'uso di rubriche valutative);
2. progettare discussioni, domande e compiti che suscitino evidenze dell'apprendimento;
3. fornire feedback per far avanzare gli studenti nell'apprendimento;
4. attivare gli studenti come proprietari del proprio apprendimento (per esempio attraverso l'autovalutazione);
5. attivare gli studenti come risorse di insegnamento e apprendimento reciproco (per esempio attraverso la valutazione tra pari) (Wiliam, 2018).

Tra queste strategie, le prime tre sono di fondamentale importanza per aiutare gli studenti nel raggiungimento di obiettivi di apprendimento specifici; mentre l'autovalutazione e la valutazione tra pari costituiscono strategie che, potenzialmente, favoriscono

in modo particolare i processi di autoregolazione dell'apprendimento³.

Più nello specifico, nel contesto dell'AfL, gli studenti sono coinvolti nel processo valutativo: conoscono gli obiettivi di apprendimento e i criteri per valutare il raggiungimento dei risultati attesi e ricevono feedback per agevolare il loro progresso. Tuttavia, questo processo è principalmente eterodiretto.

Nel caso dell'AaL, gli studenti diventano protagonisti del processo valutativo: identificano i loro obiettivi di apprendimento, gestiscono il processo per conseguirli e valutano la qualità sia dei prodotti sia dei processi coinvolti. Inoltre, si possono individuare altre caratteristiche delle strategie incentrate sull'autoregolazione e sulla sostenibilità dell'apprendimento: offrono agli studenti opportunità di apprendimento che vanno oltre il mero richiamo e utilizzo delle conoscenze pregresse; comportano una analisi sulla qualità dei prodotti e delle prestazioni. In breve, tale approccio richiede agli studenti di assumere un ruolo maggiormente attivo e riflessivo nei processi valutativi, garantendo l'attivazione di processi metacognitivi e autoregolativi.

3. Formulazione degli item e analisi della validità di contenuto

Il quadro teorico delineato ha guidato la formulazione iniziale di un pool di 20 item, valutati attraverso una scala di tipo Likert a sei (6) passi per esprimere la frequenza di utilizzo di ogni strategia, assumendo come riferimento temporale un anno scolastico (vedi *Sezione 7, infra*, per la versione finale della scala).

La scala prevede due dimensioni: *Strategie orientate al miglioramento degli apprendimenti* (S - AfL) e *Strategie orientate all'autoregolazione e alla sostenibilità dell'apprendimento* (S - AaL).

La prima dimensione (S - AfL) comprende le seguenti strategie:

- condivisione degli obiettivi di apprendimento;
- condivisione dei criteri di valutazione;
- feedback formativo, che consiste nel fornire agli studenti informazioni sui punti di forza e di debolezza e su come migliorare.

La seconda dimensione (S - AaL) include le seguenti strategie:

- prevedere l'utilizzo, da parte degli studenti, di modalità di pianificazione e di monitoraggio dell'apprendimento, ad esempio attraverso un diario/agenda di apprendimento;
- prevedere l'utilizzo, da parte degli studenti, di modalità di autovalutazione e revisione dei compiti propri e dei compagni basate su criteri e indicatori.

3 La possibilità di favorire i processi di autoregolazione dipende dalle modalità con cui si implementano le strategie di autovalutazione e di valutazione tra pari. Per esempio, è determinante se gli studenti partecipano alla definizione dei criteri e degli indicatori oppure se utilizzano quelli già stabiliti; se forniscono semplicemente un voto o un giudizio sintetico o se sono coinvolti nella revisione del proprio lavoro o di quello dei compagni.

- lavorare sui criteri di qualità, ad esempio utilizzando exemplar e creando/rielaborando compiti.

La prima versione della scala è stata valutata da un panel di otto esperti. Ciascun esperto ha compilato un protocollo di valutazione per giudicare la chiarezza e la rilevanza di ciascun item su una scala di tipo Likert a 4 passi. Gli esperti sono stati anche invitati a decidere se mantenere o rimuovere l'item dalla scala, fornendo eventuali commenti e suggerimenti per modifiche. Successivamente, sono stati calcolati il *Content Validity Index* (CVI, Lynn, 1986) e il coefficiente di Fleiss (1971), per valutare il grado di accettabilità dei singoli item. Gli item con una percentuale di accordo intergiudice inferiore al 70% e con un indice CVI inferiore a 0,86 sono stati scartati. Il processo ha portato all'eliminazione di due item, ma altrettanti sono stati aggiunti su suggerimento degli esperti.

Successivamente, la scala è stata sottoposta a un campione pilota composto da 18 docenti di scuola primaria e secondaria per verificare eventuali difficoltà tecniche nella raccolta dei dati e nell'interpretazione degli item. Dato che non sono state rilevate criticità durante la fase pilota, la scala è stata successivamente somministrata al campione di validazione.

4. Modalità di somministrazione della scala e descrizione del campione

La scala è stata somministrata in forma anonima e previa sottoscrizione del consenso informato tramite un Modulo Google, inviando un invito via email a tutti gli istituti scolastici statali italiani nel periodo compreso tra maggio e luglio 2022.

Il campione è composto da 1.545 docenti con un'età compresa tra i 22 e i 70 anni ($M = 47,92$; $DS = 10,11$). L'esperienza media degli insegnanti nel campo dell'istruzione è di 17,26 anni ($DS = 11,39$), con un range che va da zero (primo anno di insegnamento) a 46 anni. La *Tabella 1* illustra le caratteristiche dei docenti rispondenti, confrontate con quelle della popolazione target dell'anno scolastico 2020/21, estratte dal Portale Unico dei Dati della Scuola.

Nonostante la natura non probabilistica del campione, esso possiede un buon livello di rappresentatività in relazione alle caratteristiche considerate della popolazione target, ad eccezione della provenienza geografica, con una partecipazione inferiore dei docenti delle regioni del sud e delle isole.

Variabili	Popolazione docenti scuola statale Primaria e Secondaria ^a N = 806.219		Campione N = 1.545	
	n	%	n	%
Genere				
Maschio	164.971	20,5	261	16,9
Femmina	641.248	79,5	1273	82,4
Nessuno dei due/Preferisco non rispondere	-	-	11	0,7
Fascia d'età				
Fino a 34 anni	77.285	9,6	196	12,7
Tra i 35 e i 44 anni	183.890	22,8	346	22,4
Tra i 45 e i 54 anni	267.569	33,2	537	34,8
Oltre 54 anni	277.475	34,4	466	30,2
Ordine di scuola				
Primaria	292.356	36,3	560	36,2
Secondaria I grado	202.379	25,1	431	27,9
Secondaria II grado	311.484	38,6	554	35,9
Ruolo				
Titolare [†]	609.761	75,6	1200	77,7
Supplente	196.458	24,4	345	22,3
Tipologia di posto				
Comune	640.381	79,4	1330	86,1
Sostegno	165.838	20,6	215	13,9
Area geografica[‡]				
Nord ovest	200.205	24,8	437	28,3
Nord est	134.283	16,7	349	22,6
Centro	163.940	20,3	466	30,2
Sud	207.870	25,8	214	13,9
Isole	99.921	12,4	79	5,1

Tabella 1. Confronto tra le caratteristiche del campione e quelle della popolazione target. *Note:* ^(†)Elaborazioni personali su dati del Portale Unico dei Dati della Scuola (A.S. 2020/21). ^(‡)Inclusi i docenti in anno di prova. ^(§)Nel Portale non sono inclusi i dati del personale docente delle scuole delle province autonome di Aosta, Trento e Bolzano, compresi invece nel presente campione.

5. Risultati

Per verificare la struttura dimensionale della scala, sono state condotte sia un'analisi fattoriale esplorativa (EFA) che un'analisi fattoriale confermativa (CFA). Successivamente, sono stati valutati l'affidabilità degli item, la coerenza interna dei fattori e la validità convergente e discriminante dei fattori.

Le statistiche descrittive e l'EFA sono state eseguite utilizzando il software SPSS Statistics v. 28, mentre la CFA è stata condotta con il pacchetto Lavaan 0.6 – 11 per il software R (Rosseel, 2012).

Il campione è stato diviso casualmente in due sottocampioni composti da 765 e 780 soggetti per eseguire EFA e CFA su soggetti differenti. I due campioni sono stati confrontati per garantire la comparabilità

delle caratteristiche di base degli insegnanti. Sono stati utilizzati il *t*-test o il test chi-quadro per confrontare l'età, gli anni di esperienza nell'insegnamento, il genere, l'ordine di scuola, la tipologia di posto e l'area geografica. Non sono state riscontrate differenze significative, pertanto è possibile concludere che i due sottocampioni non sono significativamente diversi per quanto riguarda le caratteristiche considerate.

5.1 Analisi fattoriale esplorativa

L'EFA è stata condotta sul sottocampione composto da 765 rispondenti. La *Tabella 2* mostra le statistiche descrittive delle variabili della scala StraVI relative al primo sottocampione.

Item	M	ES	DS	Asimmetria (ES = ,088)	Curtosi (ES = ,177)
S1	4,73	,044	1,225	-,831	-,062
S2	4,44	,053	1,474	-,623	-,637
S3	4,86	,042	1,162	-,786	-,128
S4	4,67	,046	1,262	-,730	-,159
S5	4,72	,044	1,216	-,714	-,153
S6	5,24	,037	1,014	-1,333	1,361
S7	5,25	,035	,962	-1,149	,638
S8	3,69	,051	1,419	-,099	-,796
S9	2,54	,050	1,374	,681	-,259
S10	2,73	,054	1,507	,591	-,605
S11	2,69	,053	1,478	,656	-,434
S12	2,20	,049	1,342	1,097	,501
S13	4,16	,050	1,375	-,286	-,726
S14	4,50	,047	1,301	-,577	-,390
S15	3,86	,044	1,225	,015	-,487
S16	2,75	,053	1,472	,587	-,516
S17	1,96	,048	1,314	1,308	1,093
S18	3,14	,053	1,465	,210	-,760
S19	3,50	,052	1,425	,066	-,725
S20	2,19	,047	1,304	1,118	,803

Tabella 2. Statistiche descrittive della scala StraVI primo sottocampione Nota: *n* = 765. Non ci sono valori mancanti. Tutte le variabili hanno valori compresi tra 1 e 6, coprendo l'intera gamma delle opzioni di risposta

La distribuzione univariata degli item presenta valori $< |2|$, per cui non sono presenti problemi di anomalia. È stata anche verificata la distribuzione normale multivariata degli item, calcolando i coefficienti di asimmetria e curtosi multivariata di Mardia (1970). Il coefficiente di curtosi multivariata di Mardia per la presente distribuzione è pari a 515,05 ($z = 34,99$; $p = ,000$), quello di simmetria è invece pari a 38,70 ($c^2 = 4934,31$, $gdl = 1540$; $p = ,000$). Entrambi si discostano significativamente dalla normale multivariata, risultato atteso considerata la numerosità del campione.

Prima di procedere all'estrazione dei fattori, sono stati verificati i requisiti necessari per l'applicazione dell'EFA. Sono stati condotti il test di adeguatezza campionaria di Kaiser-Meyer-Olkin (KMO; Kaiser, 1970) e il test di Sfericità di Bartlett (Bartlett, 1954), oltre a verificare che il Determinante fosse diverso da zero.

L'analisi del KMO ha mostrato un indice di 0,911 quindi nella fascia dei valori eccellenti (Kaiser & Rice, 1974). Il test di Sfericità di Bartlett (BTS) ($\chi^2 = 8362,234$; $gdl = 190$; $p = ,000$) indica una significativa correlazione tra le variabili. Infine, il Determinante è diverso da zero, dunque non ci sono combinazioni lineari perfette tra variabili. È stato inoltre calcolato il valore del KMO relativo ad ogni singola variabile (MSA; *Measure of Sampling Adequacy*): tutte le variabili presentano un indice MSA superiore a 0,80, soddisfacendo la soglia di accettabilità.

L'EFA è stata condotta utilizzando il metodo di estrazione dei minimi quadrati non ponderati, con una rotazione obliqua (Oblimin con normalizzazione Kaiser), quest'ultima in considerazione della presunta correlazione tra i fattori.

Per determinare il numero di fattori da estrarre, è stato utilizzato il metodo delle analisi parallele (Horn,

1965), il quale ha suggerito inizialmente la presenza di cinque fattori.

I passaggi necessari per la ridefinizione del modello sono stati iterativi. In sintesi, il processo ha portato all'eliminazione dei seguenti item per garantire una struttura semplice (Thurstone, 1947): S5, S8, S9, S12, S13, S14, S15.

L'ultimo modello estratto, con quattro fattori e 13 item, spiega il 62,2% della varianza (MKO = ,867; BTS = 5123,044; $gdl = 78$; $p = ,000$). La struttura che emerge è molto semplice e suggerisce una definizione precisa dei costrutti (vedi Tabella 6).

Al termine della definizione del modello, gli item sono stati rinominati per favorirne l'identificazione al fattore.

A livello teorico, si ipotizza che i fattori *Condivisione e chiarimento di oda e criteri valutativi* (CC) e *Feedback formativo* (FF) siano riflessivi di una dimensione di ordine superiore, ovvero *Strategie valutative orientate al miglioramento degli apprendimenti*. Analogamente, si ritiene che i fattori *Supporto all'autoregolazione dell'apprendimento* (SAA) e *Supporto allo*

sviluppo del giudizio valutativo (SGV) siano riflessivi della dimensione *Strategie valutative orientate all'autoregolazione e alla sostenibilità dell'apprendimento*. L'ipotesi della struttura fattoriale gerarchica sarà testata nell'analisi fattoriale confermativa.

5.2 Analisi fattoriale confermativa

Il modello emerso dall'analisi fattoriale esplorativa è stato sottoposto ad analisi fattoriale confermativa utilizzando il secondo sottocampione.

La Tabella 3 mostra le statistiche descrittive della scala per il secondo sottocampione. Anche in questo caso, si osserva una distribuzione univariata degli item che non si discosta in modo sostanziale da quella normale. Tuttavia, gli indici di curtosi e asimmetria multivariati risultano significativi: l'indice di asimmetria multivariata di Mardia è pari a 16,292 ($c^2 = 2117,92$; $gdl = 455$; $p = ,000$), mentre l'indice di curtosi multivariata di Mardia è pari a 218,80 ($z = 16,83$; $p = ,000$).

Item	M	ES	DS	Asimmetria (ES = ,088)	Curtosi (ES = ,175)
CC1	4,66	,042	1,168	-,526	-,554
CC2	4,35	,051	1,417	-,520	-,711
CC3	4,85	,043	1,189	-,802	-,103
CC4	4,61	,047	1,313	-,655	-,415
FF1	5,25	,035	,977	-1,279	1,270
FF2	5,25	,033	,933	,569	2,680
SAA1	2,68	,052	1,459	,573	-,581
SAA2	2,67	,053	1,467	,636	-,529
SGV1	2,65	,053	1,491	,606	-,602
SGV2	2,01	,046	1,342	1,342	1,191
SGV3	2,98	,051	1,420	,331	-,631
SGV4	3,38	,049	1,360	,093	-,568
SGV5	2,16	,046	1,294	1,089	,648

Tabella 3. Statistiche descrittive della scala StraVI secondo sottocampione. Nota: $n = 780$. Non ci sono valori mancanti. Tutte le variabili hanno valori compresi tra 1 e 6 (ad eccezione di FF2 il cui valore minimo è 2).

Per la stima dei parametri è stato utilizzato un metodo di estrazione robusto MLM (*Maximum likelihood estimation with robust standard errors and a Satorra-Bentler scaled test statistic*) che utilizza il chi quadro con la correzione di Satorra-Bentler (S-Bc²; Satorra & Bentler, 2001).

Per valutare la bontà di adattamento del modello, sono stati utilizzati i seguenti indici con i relativi cutoff suggeriti da Kline (2016):

- *Model chi-square*: testa l'ipotesi nulla che il modello si adatti esattamente ai dati osservati.
- *Comparative Fit Index* (CFI; Bentler, 1990): indice di adattamento detto "incrementale" o "relativo" perché confronta la discrepanza del modello rispetto a un modello ideale; valori superiori a 0,90 o, più restrittivo, a 0,95 indicano un buon adattamento.
- *Tucker-Lewis Index* (TLI; Tucker & Lewis, 1973): altro indice di adattamento incrementale; valori maggiori di 0,90 indicano un buon adattamento.
- *Root Mean Square Error Approximation* (RMSEA; Steiger, 1990; Steiger & Lind, 1980): indice assoluto (confronta il modello rispetto al fit perfetto ai dati osservati), i criteri di cutoff sono i seguenti: " 0,05 (buono), tra 0,05 e 0,08 (accettabile), $\geq 0,10$ (scarso).
- *Standardized Root Mean Square Residual* (SRMR): altro indice assoluto, con valori accettabili " 0,08 (Hu & Bentler, 1999).

Per l'identificazione dei fattori è stato sempre utilizzato il "market method", che fissa il primo *loading* di ogni fattore a 1 e stima liberamente la varianza di ogni fattore e gli altri parametri delle saturazioni fattoriali (*factor loadings*), così come le covarianze residue.

È stato specificato un modello di misurazione con quattro fattori di primo ordine per poi confrontarlo con un possibile modello gerarchico, annidato al primo, che riflette al meglio il framework teorico.

In primo luogo, è opportuno evidenziare come il modello non presenta particolari criticità: le stime di

carico standardizzato sono tutte superiori a 0,5 e risultano essere significative; non ci sono varianze negative; non ci sono correlazioni o coefficienti strutturali standardizzati maggiori di |1|. Analizzando gli indici di fit, si osserva che l'adattamento del modello è buono (Tabella 4), nonostante il χ^2 del modello sia significativo, come prevedibile data l'ampia dimensione del campione.

Successivamente, è stato testato il modello gerarchico, maggiormente in linea con il framework teorico e supportato anche dall'alta correlazione tra i fattori di primo ordine.

Il modello gerarchico vede il fattore *Strategie valutative orientate al miglioramento degli apprendimenti* (S - AfL) riflessivo dei fattori di primo ordine CC e FF e il fattore *Strategie valutative orientate all'auto-regolazione e alla sostenibilità dell'apprendimento* (S - AaL) riflessivo dei fattori SAA e SGV. Anche questo secondo modello si adatta bene ai dati (Tabella 4).

Modello	S-Bc ²	gdl	RMSEA*	CFI*	TLI*	SRMR
Modello 1 1° ordine a 4 fattori Num. parametri: 32	218,873***	59	,062 (90% CI [,053; ,071])	,963	,952	,047
Modello 2 2° ordine 2 fattori di 2° ordine Num. parametri: 31	223,832***	60	,062 (90% CI [,054; ,071])	,963	,951	,048

Tabella 4. Indici di fit dei modelli fattoriali sottoposti a CFA della scala StraVI. Nota: n = 780. Nessun dato mancante.

RMSEA = Root Mean Square Error Approximation;
CFI = Comparative Fit Index; TLI = Tucker-Lewis Index;
SRMR = Standardized Root Mean Square Residual. (*) Versione robusta degli indici. (***) $p < ,001$.

Per confrontare il modello gerarchico (Modello 2) con il modello di primo ordine (Modello 1), è stato utilizzato il test della differenza del Chi quadro aggiustata per la media di Satorra-Bentler (Satorra & Bentler, 2001). Tuttavia, i test di differenza del Chi quadro possono essere influenzati dalla dimensione del campione, rendendo significative anche differenze di

adattamento di piccola entità (Kline, 2016; Schermelleh-Engel et al., 2003). Pertanto, per valutare l'invarianza della bontà di adattamento, è stata utilizzata la differenza degli indici CFI. Secondo Cheung e Rensvold (2002) un CFI minore o uguale a 0,01 indica una mancanza sostanziale di differenza di adattamento tra due modelli annidati; secondo Meade e colleghi (2008), un cutoff di 0,002, è valido in un'ampia gamma di condizioni.

La differenza tra χ^2 dei modelli 1 e 2 è statisticamente significativa ($S-B^2_{(1)} = 5,475, p = ,019$), ma CFI = 0. Pertanto, considerando il valore di CFI, si può escludere una differenza sostanziale tra gli adattamenti dei due modelli. Si preferisce quindi il modello gerarchico, che risulta più parsimonioso con meno parametri e più gradi di libertà (Pavlov et al., 2020), oltre a essere giustificato dal quadro teorico di riferimento.

Il modello gerarchico implica che i fattori di primo ordine riflettono una dimensione superiore piuttosto che riferirsi a costrutti concettualmente separati. Come evidenziato nella Tabella 5, i due costrutti di secondo ordine mostrano una correlazione moderata.

La Figura 1 illustra il modello fattoriale della scala, mentre la Tabella 6 mostra le saturazioni fattoriali del modello finale.

	CC	FF	SAA	SGV	SAfL	SAaL
CC	–					
FF	,630	–				
SAA	,497	,330	–			
SGV	,453	,301	,805	–		
SAfL	,974	,647	,510	,465	–	
SAaL	,529	,352	,939	,856	,543	–

Tabella 5. Matrice di correlazione dei fattori latenti della scala CoVI. Nota: tutte le correlazioni sono significative a livello $p < ,001$.

Tabella 6. Saturazioni fattoriali per EFA e CFA della scala StraVI. Nota: In corsivo gli item; in **grassetto** le saturazioni primarie dell'EFA; S - AfL e S - AaL sono fattori di secondo ordine; Ind.r. = Individual reliability.

	EFA campione 1 (n = 765)					CFA campione 2 (n = 780)						
	CC	FF	SAA	SGV	<i>h</i> ²	CC	FF	SAA	SGV	S-AfL	S-AaL	Ind.r.
CC1	,729	,038	,045	,092	,606	,719						,517
CC2	,866	–,096	,016	–,009	,661	,715						,511
CC3	,709	,134	–,083	–,056	,622	,760						,578
CC4	,799	,023	–,051	–,003	,685	,855						,731
FF1	,070	,761	,014	,037	,655		,824					,679
FF2	–,030	,915	–,081	–,004	,833		,819					,671
SAA1	,104	,041	–,549	,221	,597			,833				,694
SAA2	,056	,088	–,829	,042	,805			,840				,706
SGV1	,017	,018	–,196	,640	,633				,797			,635
SGV2	–,029	–,091	–,258	,618	,609				,814			,663
SGV3	,170	,042	,042	,575	,436				,650			,423

SGV4	-,009	,110	,110	,717	,486				,627			,393
SGV5	-,017	-,130	-,130	,614	,461				,732			,536
CC										,974		
FF										,647		
SAA											,939	
SGV												,856

Tabella 6. Saturazioni fattoriali per EFA e CFA della scala StraVI. Nota: In corsivo gli item; in grassetto le saturazioni primarie dell'EFA; S - AfL e S - AaL sono fattori di secondo ordine; Ind.r. = Individual reliability.

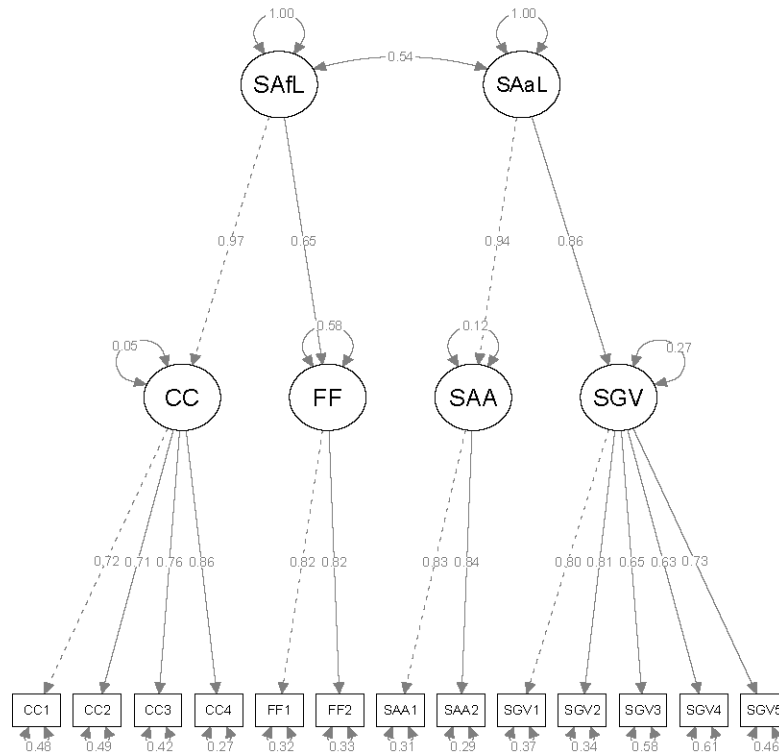


Figura 1. Modello fattoriale della scala StraVI (soluzione standardizzata).

5.3 Affidabilità degli item, coerenza interna dei fattori e validità convergente e discriminante

A partire dalle stime standardizzate delle saturazioni fattoriali e dei termini di errore ottenuti tramite l'analisi fattoriale confermativa, sono stati calcolati l'affidabilità dei singoli item, la coerenza interna dei fattori di primo e secondo ordine e la loro validità convergente e discriminante.

L'affidabilità dei singoli item che saturano i fattori è stata valutata mediante l'indice di attendibilità del singolo indicatore di un fattore (*single item o individual reliability*). I risultati mostrano indici di affidabilità maggiori di 0,30 per tutti gli item, mostrando una buona affidabilità (Tabella 6).

Per quanto riguarda l'affidabilità dei fattori, è stato utilizzato l'indice di attendibilità del costrutto (*com-*

posite o construct reliability, CR) come misura della coerenza interna. L'affidabilità dei fattori di secondo ordine (S - AfL e S - AaL) è stata valutata attraverso il coefficiente omega a livello 1 (proporzione di varianza totale dei punteggi osservati dovuta al fattore di secondo ordine), il coefficiente omega a livello 2 (proporzione della varianza totale dei fattori di primo ordine spiegata dalla presenza del fattore di secondo ordine) e il coefficiente omega parziale a livello 1 (proporzione della varianza dei punteggi osservati dovuta al fattore di secondo ordine dopo l'eliminazione dell'effetto di unicità dei fattori di primo ordine).

Secondo la regola empirica, una stima pari o superiore a 0,70 indica una buona affidabilità (Hair et al., 2014). Come indicato nella Tabella 7, tutti i fattori hanno una buona affidabilità.

Fattore	α	CR	AVE	α	α L1	α L2	α parziale L1
CC	,844	,848	,584	,848			
FF	,805	,806	,675	,806			
SAA	,824	,823	,700	,824			
SGV	,845	,848	,530	,847			
S α - α AfL	,849	,806	,684		,807	,813	,875
S α - α AaL	,880	,893	,807		,768	,895	,882

Tabella 7 Statistiche di affidabilità e validità convergente dei fattori della scala StraVI. Nota: CR = Construct Reliability; AVE = Average Variance Extracted.

Per valutare la validità convergente dei fattori, è stato considerato innanzitutto il peso fattoriale dei singoli item, assumendo come parametro di accettabilità un *factor loading* maggiore di 0,50 (Hair et al., 2014; Kline, 2016). Inoltre, l'indice AVE (*Average Variance Extracted*) dovrebbe essere uguale o maggiore di 0,50 e minore dell'indice CR (Fornell & Larcker, 1981). Tutti gli item hanno un *factor loading* maggiore di 0,50, inoltre ogni fattore presenta un indice AVE maggiore di 0,50 e minore dell'indice CR, indicando una buona validità convergente.

Per quanto riguarda la validità discriminante, è

stata verificata attraverso l'applicazione del criterio di Fornell-Larcker (1981), secondo il quale la validità discriminante tra due fattori è garantita se l'indice AVE di ciascun fattore è maggiore del quadrato della loro correlazione (Fornell & Larcker, 1981). Ciascun fattore di primo ordine deve mostrare validità discriminate rispetto a tutti gli altri fattori del modello, ad eccezione dei fattori di ordine superiore di cui fanno parte, rispetto ai quali non è richiesta. Stesso principio vale per i fattori di ordine superiore (Sarstedt et al., 2019).

Tutti i fattori del modello mostrano di possedere validità discriminante (Tabella 8).

	CC	FF	SAA	SGV	SAfL	SAaL
CC	,584					
FF	,397	,675				
SAA	,247	,109	,700			
SGV	,205	,091	,648	,530		
SAfL	,949	,419	,260	,216	,684	
SAaL	,280	,124	,882	,733	,295	,807

Tabella 8. Validità discriminante dei fattori della scala StraVI. Nota: I valori sulla diagonale sono i valori dell'AVE (*Average Variance Extracted*); gli elementi fuori dalla diagonale sono i quadrati delle correlazioni tratte dalla matrice di correlazione dei fattori latenti.

6. Discussione e conclusioni

Lo studio qui presentato si è proposto di sviluppare e validare una scala per rilevare la frequenza di utilizzo di strategie valutative di tipo formativo da parte dei docenti nell'ambito dell'istruzione primaria e secondaria in Italia. Considerata la mancanza di strumenti adeguati a una valutazione specifica delle strategie riconducibili agli approcci di AfL e AaL, la scala StraVI è stata sviluppata per rilevare le strategie connesse a tali approcci, considerando anche la valutazione sostenibile come aspetto dell'AaL.

Lo strumento è stato validato su un campione di 1.545 docenti in servizio presso scuole statali primarie e secondarie di primo e secondo grado, con una buona rappresentatività della popolazione target, ad eccezione della provenienza geografica, per cui si è riscontrata una sottorappresentazione dei docenti delle scuole del sud e delle isole rispetto alla popolazione di riferimento. Tuttavia, è importante sottoli-

neare che il campionamento non è stato probabilistico, costituendo un limite dello studio. Inoltre, va segnalata la mancanza di un confronto con una misura esterna che potesse attestare la validità predittiva dello strumento.

Nella costruzione della scala, si è distinto tra strategie direttamente connesse all'approccio di AfL, come la condivisione degli obiettivi di apprendimento e dei criteri di valutazione e il feedback dell'insegnante, e altre più strettamente legate all'approccio di AaL e di *sustainable assessment*, che favoriscono maggiormente l'attivazione di processi di autoregolazione dell'apprendimento e l'acquisizione del giudizio valutativo.

L'analisi fattoriale ha confermato la struttura multidimensionale della scala, composta da 13 item, evidenziando la presenza di quattro fattori di primo ordine: *Condivisione e chiarimento degli obiettivi e criteri valutativi* (CC), *Feedback formativo* (FF), *Supporto all'autoregolazione dell'apprendimento* (SAA),

e *Supporto allo sviluppo del giudizio valutativo (SGV)*. Questi fattori, a loro volta, riflettono due fattori di secondo ordine: *Strategie valutative orientate al miglioramento degli apprendimenti (S - AfL)* e *Strategie valutative orientate all'autoregolazione e alla sostenibilità dell'apprendimento (S - AaL)*.

Le analisi di attendibilità e validità hanno confermato la solidità del modello, evidenziando buone stime di affidabilità sia per gli item che per i fattori, oltre a una validità convergente e discriminante.

In conclusione, la scala StraVI colma una lacuna nell'ambito degli strumenti disponibili, poiché è in grado di mettere a fuoco le specificità delle strategie valutative di tipo formativo riconducibili a diversi approcci valutativi, compresi l'AaL e il *sustainable assessment*, meno esplorati empiricamente. Questo strumento può essere utilizzato per fornire un quadro chiaro delle strategie valutative di tipo formativo utilizzate dai docenti. Ulteriori ricerche potrebbero approfondire l'efficacia di tali strategie nel migliorare i risultati di apprendimento degli studenti e nel promuoverne i processi metacognitivi e di autoregola-

zione dell'apprendimento, oltre a esplorare la relazione tra le strategie valutative utilizzate e altri aspetti della pratica valutativa e didattica dei docenti.

7. Appendice: Scala delle Strategie Valutative degli insegnanti (StraVI)

7.1 Istruzioni

Di seguito troverà un elenco di strategie valutative che possono essere utilizzare in classe. Pensando alla sua pratica didattica in relazione a una sua classe tipo, indichi la frequenza con cui ogni situazione si verifica nel corso di un anno scolastico.

Nel corso di un anno scolastico...

Formato di risposta: 1 = *mai*; 2 = *raramente*; 3 = *qualche volta*; 4 = *di frequente*; 5 = *quasi sempre*; 6 = *sempre*.

ID	Item
CC1	Condivido con gli studenti gli obiettivi di apprendimento collegati a un compito o a un'attività.
CC2	Condivido con gli studenti i criteri e gli indicatori che saranno utilizzati per valutare i loro prodotti o prestazioni (per es. attraverso la condivisione di rubriche).
CC3	Mi assicuro che gli studenti abbiano compreso gli obiettivi di apprendimento prefissati e, se necessario, intervengo.
CC4	Mi assicuro che gli studenti abbiano compreso i criteri e gli indicatori con cui verranno valutati i loro prodotti o prestazioni e, se necessario, intervengo.
FF1	Dopo un compito svolto o un'attività, fornisco agli studenti feedback riguardanti punti di forza e di debolezza delle loro prestazioni o prodotti.
FF2	Dopo un compito svolto o un'attività, fornisco agli studenti indicazioni su cosa poter fare per migliorare il prodotto o la prestazione (per es. indicando quali contenuti è necessario approfondire).
SAA1	All'inizio di un percorso di apprendimento, chiedo agli studenti di provare a pianificare e organizzare le proprie azioni e strategie per il raggiungimento degli obiettivi (ad es. attraverso un diario/agenda di apprendimento).
SAA2	Al termine di un percorso di apprendimento, chiedo agli studenti di valutare le azioni intraprese, le strategie utilizzate e il livello di raggiungimento degli obiettivi (ad es. attraverso un portfolio o una biografia cognitiva).
SGV1	Dopo una prima consegna di un compito, chiedo agli studenti di produrre una revisione (non solo un voto o un giudizio sintetico) del PROPRIO lavoro sulla base di criteri e indicatori DEFINITI INSIEME.
SGV2	Dopo una prima consegna di un compito, chiedo agli studenti di produrre una revisione (non solo un voto o un giudizio sintetico) del lavoro DEI COMPAGNI sulla base di criteri e indicatori DEFINITI INSIEME.
SGV3	Propongo agli studenti esempi o buoni modelli di ciò che ci si aspetta da loro e ne discuto gli indicatori di qualità (cosa fa di quell'esempio un "buon" modello).
SGV4	Chiedo agli studenti di utilizzare il feedback ricevuto da me o dai compagni per rielaborare il proprio lavoro.
SGV5	Chiedo agli studenti di creare prove di valutazione (es. domande a risposta multipla su un argomento studiato) per sé stessi e/o per i propri compagni.

Riferimenti bibliografici

- Assessment Reform Group (ARG) (2002). *Assessment is for learning: 10 principles. Research-based principles to guide classroom practice*. https://assessmentreformgroup.files.wordpress.com/2012/01/10principles_english.pdf
- Bartlett, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 296–298.
- Batini, F., & Guerra, M. (2020). Gli effetti della valutazione formativa sull'apprendimento nella scuola primaria. Una revisione sistematica. *Pedagogia più Didattica*, 6(2), 78–93.
- Ben-Eliyahu, A. (2021). Sustainable learning in education. *Sustainability*, 13(8), Article 4250. <https://doi.org/10.3390/su13084250>
- Bennett, R. E. (2011). Formative assessment: A critical review. *Assessment in Education: Principles, Policy & Practice*, 18(1), 5–25. <https://doi.org/10.1080/0969594X.2010.513678>
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, 107, 238–246. <https://doi.org/10.1037/0033-2909.107.2.238>
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in Continuing Education*, 22(2), 151–167. <https://doi.org/10.1080/713695728>
- Boud, D. (2022). Assessment-as-learning for the development of students' evaluative judgment. In Z. Yan, & L. Yang (Eds.), *Assessment as learning: Maximising opportunities for student learning and achievement* (pp. 25–37). Routledge. [Versione Kindle MAC].
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment & Evaluation in Higher Education*, 41(3), 400–413. <https://doi.org/10.1080/02602938.2015.1018133>
- Cheung, G. W., & Resnold, R. B. (2002). Evaluating Goodness of fit indexes for testing measurement invariance. *Structural Equation Modelling: A Multidisciplinary Journal*, 9, 233–255. https://doi.org/10.1207/S15328007SEM0902_5
- Clark, I. (2012). Formative assessment: Assessment is for self-regulated learning. *Educational Psychology Review*, 24(2), 205–249. <https://doi.org/10.1007/s10648-011-9191-6>
- Corsini, C. (2023). *La valutazione che educa. Liberare insegnamento e apprendimento dalla tirannia del voto*. FrancoAngeli.
- Crooks, T. (2011). Assessment for learning in the accountability era: New Zealand. *Studies in Educational Evaluation*, 37(1), 71–77. <https://doi.org/10.1016/j.stueduc.2011.03.002>
- Earl, L. M. (2013). *Assessment as Learning: Using classroom assessment to maximize student learning* (2nd ed.). Corwin.
- Fleiss, J. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76, 378–382. <https://doi.org/10.1037/h0031619>
- Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research (JMR)*, 18(1), 39–50. <http://www.jstor.org/stable/3151312>
- Galliani, L. (2009). *Web ontology della valutazione educativa*. Pensa MultiMedia.
- Graham, L., Berman, J., & Bellert, A. (2015). *Sustainable learning*. Cambridge University Press.
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (7th ed.). Person Education Limited.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses on achievement*. Routledge.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179–185. <https://doi.org/10.1007/BF02289447>
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1), 1–55.
- Kaiser, H. F. (1970). A second-generation little jiffy. *Psychometrika*, 35(4), 401–415. <https://doi.org/10.1007/BF02291817>
- Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark 4. *Educational and Psychological Measurement*, 34(1), 111–117. <https://doi.org/10.1177/001316447403400115>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
- Lam, R. (2019). Teacher assessment literacy: Surveying knowledge, conceptions and practices of classroom-based writing assessment in Hong Kong. *System*, 81, 78–89. <https://doi.org/10.1016/j.system.2019.01.006>
- Lynn, M. R. (1986). Determination and quantification of content validity. *Nursing Research*, 35(6), 382–385. <https://doi.org/10.1097/00006199-198611000-00017>
- Mardia, K. V. (1970). Measures of multivariate Skewness and Kurtosis with applications. *Biometrika*, 57(3), 519–530. <https://doi.org/10.2307/2334770>
- Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, 93(3), 568–592. <https://doi.org/10.1037/0021-9010.93.3.568>
- Meusen-Beekman, K., Joosten-ten Brinke, D., & Boshuizen, E. (2016b). De retentie van zelfregulatie, motivatie en self-efficacy in het voortgezet onderwijs na formatieve assessments in het basisonderwijs. *Pedagogische Studiën*, 93(3), 136–153.
- Panadero, E., Alonso-Tapia, J., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences*, 22(6), 806–813. <https://doi.org/10.1016/j.lindif.2012.04.007>
- Pavlov, G., Shi, D., & Maydeu-Olivares, a. (2020) Chi-square Difference Tests for Comparing Nested Models: An Evaluation with Non-normal Data, *Structural Equation Modelling: A Multidisciplinary Journal*, 27(6), 908–917. <https://doi.org/10.1080/10705511.2020.1717957>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1–36. <https://doi.org/10.18637/jss.v048.i02>
- Sarstedt, M., Hair Jr, J. F., Cheah, J. H., Becker, J. M., & Ringle, C. M. (2019). How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australasian Marketing Journal (AMJ)*, 27(3), 197–211. <https://doi.org/10.1016/j.ausmj.2019.05.003>
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, 66(4), 507–514. <https://doi.org/10.1007/BF02296192>
- Schellekens, L. H., Bok, H. G., de Jong, L. H., van der Schaaf, M. F., Kremer, W. D., & van der Vleuten, C. P. (2021). A scoping review on the notions of Assessment as Learning (AaL), Assessment for Learning (AfL), and Assessment of Learning (AoL). *Studies in Educational Evaluation*, 71, Article 101094. <https://doi.org/10.1016/j.stueduc.2021.101094>
- Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of psychological research online*, 8(2), 23–74.
- Schneider, M., & Preckel, F. (2017). Variables associated with achievement in higher education: A systematic review of meta-analyses. *Psychological Bulletin*, 143(6), 565–600. <http://dx.doi.org/10.1037/bul0000098>
- Scierri, I. D. M. (2021). Strategie e strumenti di valutazione formativa per promuovere l'apprendimento autoregolato: una rassegna ragionata delle ricerche empiriche. *Journal of Educational, Cultural and Psychological Studies*, 24, 213–227. <https://doi.org/10.7358/ecps-2021-024-scie>
- Scierri, I. D. M. (2022). Finalità e modi della valutazione: primi

- risultati di un'indagine su concezioni e strategie valutative dei docenti delle scuole italiane. In V. Boffo, & F. Togni (Eds.), *Esercizi di ricerca. Dottorato e politiche della formazione* (pp. 197–206). Firenze University Press.
- Scierri, I. D. M. (2023). Per una valutazione centrata sull'allievo: framework teorico e primi risultati di un'indagine su concezioni e strategie valutative degli insegnanti. *Lifelong Lifewide Learning*, 12(42), 83–101. <https://doi.org/10.19241/lll.v19i42.754>
- Scierri, I. D. M., Viola, M., & Capperucci, D. (2023). Gli effetti di una valutazione come apprendimento sullo sviluppo del giudizio valutativo e sull'autoefficacia degli studenti: una esperienza nella scuola primaria. *QTimes – webmagazine*, XV, 4, 290–305.
- Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research*, 25, 173–180. https://doi.org/10.1207/s15327906mbr2502_4
- Steiger, J. H., & Lind, J. C. (1980). *Statistically based tests for the number of common factors, paper presented at the annual meeting of the Psychometric Society*. Iowa City.
- Thurstone, L. L. (1947). *Multiple factor analysis*. University of Chicago Press.
- Tomlinson, C. A. (2007). Learning to love assessment. *Educational Leadership*, 65(4), 8–13.
- Trincherò, R. (2017). Attivare cognitivamente con la valutazione formante. In A. M. Notti (Ed.), *La funzione educativa della valutazione. Teorie e pratiche della valutazione educativa* (pp. 73–90). Pensa MultiMedia.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1–10. <https://doi.org/10.1007/BF02291170>
- Vertecchi, B. (2023, febbraio). Ipotesi per un esperimento. *Tuttoscuola*, XLVIII, 26–27.
- Wiliam, D. (2018). *Embedded formative assessment* (2nd ed.). Solution Tree Press.
- Wiliam, D., & Black, P. J. (1996). Meanings and consequences: a basis for distinguishing formative and summative functions of assessment? *British Educational Research Journal*, 22, 537–548.
- Wisniewski, B., Zierer, K., & Hattie, J. (2020). The power of feedback revisited: A meta-analysis of educational feedback research. *Frontiers in Psychology*, 10, Article 3087. <https://doi.org/10.3389/fpsyg.2019.03087>
- Yan, Z., & Boud, D. (2022). Conceptualising assessment-as-learning. In Z. Yan, & L. Yang, (Eds.), *Assessment as Learning. Maximising opportunities for student learning and achievement* (pp. 11–24). Routledge. [Versione Kindle MAC].