

Article

Narrative Agency and Affective Valence in Prospective Teachers' Pedagogical Autobiographies: A Computational Text Analysis of Two Cohorts

Alice Roffi and Gabriele Biagini * 

Dipartimento di Formazione, Lingue, Intercultura, Letterature e Psicologia, University of Florence,
50121 Florence, Italy; alice.roffi@unifi.it

* Correspondence: gabriele.biagini@unifi.it

Abstract

Reflective practice is central to contemporary teacher education, and pedagogical autobiographies are widely used as tools for reflection; yet the structural properties of these narratives, and their relationship to the reflective process they are meant to support, remain under-examined. This study applies computational text analysis to 361 autobiographical accounts written by prospective future Italian teachers across two consecutive cohorts (2024–25; 2025–26). The corpus covers eight pedagogical dimensions and is analysed through two rule-based indicators, Student Agency Ratio (SAR) and Sentiment Valence, both validated against human coding on a subsample, using methods that account for the nesting of dimensions within participants. Results show strong dimension-level differences in both SAR and valence, with agency and valence profiles that replicate across the two cohorts, whereas lexical richness shows a cohort-level shift. Assessment stands apart from the other dimensions: compared with them, it combines markedly more negative valence (a large within-participant effect, $d_z = 2.31$), more frequent negative or unresolved endings, lower lexical richness, and predominant anchoring in upper secondary school. By contrast, STEM and support-related memories are more often anchored in primary school and are generally more positive. These findings suggest that prospective teachers enter initial training with domain-specific narrative schemas that may shape how they interpret key areas of pedagogical practice, especially assessment. As the data come from a single institution, the findings should be read as exploratory and warrant replication with larger, multi-institution samples.

Keywords: teacher education; pedagogical autobiography; narrative agency; computational text analysis; assessment; pre-service teachers; Italian



Academic Editors: Rui Araújo and
Mariana Amaral-da-Cunha

Received: 21 April 2026

Revised: 17 June 2026

Accepted: 22 June 2026

Published: 26 June 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

The use of autobiographical writing in pre-service teacher education has a long theoretical lineage, rooted in [Connelly and Clandinin's \(1990\)](#) work on teachers' personal practical knowledge and the recognition that belief formation precedes and shapes professional practice ([Pajares, 1992](#)). Prospective teachers do not enter initial training as blank slates: they arrive with approximately 13,000 h of 'apprenticeship of observation' ([Lortie, 1975](#))—accumulated memories of teachers, classrooms, successes, and injustices that silently pre-configure their pedagogical dispositions before any formal instruction begins. This formative role is not a matter of passive imprinting, however: recent autobiographical-memory research shows that prospective teachers construct their schooling memories

agentively, selectively encoding and recalling experiences in ways congruent with their emerging professional goals (Crowe & McGarr, 2022).

Pedagogical autobiographies are designed to make these tacit schemas explicit and, ideally, open to critical revision. A student-teacher asked to narrate a memorable episode of classroom management, or an experience of assessment that felt unjust, is prompted to surface assumptions that would otherwise remain invisible. The genre is well-established in Italian teacher education (Mortari, 2009; Zannini, 2005) and has analogues in Anglo-American ‘critical incident’ frameworks (Tripp, 1993) and narrative inquiry traditions (Clandinin, 2013). In all of these traditions the autobiography is, at root, an instrument of reflective practice: it asks the prospective teacher to return to lived experience, interrogate it, and reconstruct its meaning, in the sense developed by Dewey (1933) and Schön (1983). What autobiographical writing cannot by itself guarantee, however, is the depth of that reflection (Hatton & Smith, 1995); whether the writing remains merely descriptive or becomes genuinely critical depends in part on the narrative resources the writer brings to each domain of experience—the very resources this study sets out to map.

Despite this widespread use, the structural and affective properties of pedagogical autobiographies as a genre remain largely unexamined at scale, and the few computational studies that do exist have analysed small samples of in-service teachers rather than prospective teachers in large cohorts (e.g., Barbosa & Marín-Suelves, 2024). Existing research has relied primarily on qualitative, hermeneutic analysis of small samples (Achinstein & Ogawa, 2006; Flores & Day, 2006). Two questions in particular have received little systematic attention.

First, who is the grammatical and semantic agent in prospective teachers’ pedagogical memories: the students themselves or the teacher-figures they describe? Agency allocation in autobiographical narrative has been theorised as a marker of identity positioning (Bamberg, 1997; Wortham, 2001): narrators who consistently position themselves as acted-upon rather than acting may enter teaching with a more passive professional schema.

Second, what is the affective valence of those memories, and does it vary systematically across pedagogical domains? If prospective teachers carry predominantly negative memories of assessment and predominantly positive memories of STEM learning, those affective asymmetries have direct implications for professional identity formation (Beauchamp & Thomas, 2009) and for the design of teacher education curricula.

The present study addresses both questions through computational analysis of a corpus of $N = 361$ pedagogical autobiographies produced by prospective future Italian teachers over two cohorts. The corpus is structured around eight thematic dimensions, allowing within-subject cross-domain comparison. We introduce two operationalisable constructs—Student Agency Ratio (SAR) and Sentiment Valence—computed through rule-based Italian linguistic analysis.

We further examine narrative resolution (positive, negative, mixed), lexical richness (Moving-Average Type-Token Ratio, MATTR), and school-level anchoring (primary, lower secondary, upper secondary).

The study makes three contributions to the literature. Methodologically, it demonstrates that rule-based linguistic analysis of Italian educational narratives, anchored in established sentiment resources, yields theoretically interpretable results at a scale inaccessible to qualitative methods. Empirically, it documents the Assessment dimension as a structurally distinct narrative domain—affectively negative, lexically stereotyped, and anchored in late secondary experience—with implications for how initial training addresses evaluation competence. Theoretically, it advances the concept of narrative schema as a dimension-level cognitive structure that prospective teachers bring to teacher education, and specifies how such schemas condition the reflective process that pedagogical autobiog-

raphy is intended to initiate, building on Schön's (1983) account of reflection on professional experience, Beauchamp and Thomas's (2009) account of identity formation, and Pajares's (1992) analysis of teacher belief entrenchment.

2. Theoretical Background

2.1. Pedagogical Autobiography, Reflective Practice, and Belief Formation

Reflective practice provides the theoretical rationale for using autobiography in teacher preparation. Dewey (1933) characterised reflective thought as the active and persistent reconsideration of experience in the light of the grounds that support it, and Schön (1983) recast professional learning as reflection on—and in—the situated knowledge that practice generates. Contemporary models in teacher education extend this lineage, distinguishing levels of reflection that range from mere description to critical, assumption-questioning reflection (Hatton & Smith, 1995; Larrivee, 2000) and linking the deepest forms to the transformation of entrenched frames of reference (Mezirow, 1991). Within this tradition, writing a pedagogical autobiography is not simply recollection but a structured occasion for reflection: the prospective teacher is asked to surface tacit beliefs and subject them to scrutiny. The present study takes a prior step, asking what these autobiographical narratives are structurally like before such reflective work begins, on the premise that the schema a writer brings to a domain both enables and constrains the reflection that can subsequently follow.

Teacher beliefs are notoriously resistant to change. Pajares (1992) argued that beliefs formed early in schooling when students are emotionally receptive and cognitively dependent are encoded with an affective intensity that formal university instruction rarely overcomes. Richardson (1996) extended this argument, showing that prior schooling experiences consistently outweigh coursework as predictors of beginning teachers' classroom practice. Autobiographical writing, in this literature, functions not as a cathartic exercise but as a diagnostic and potentially disruptive one: making the belief visible is a precondition for subjecting it to rational scrutiny (Korthagen & Vasalos, 2005).

Narrative identity theory (Ricoeur, 1991; McAdams, 1993) provides a complementary lens: the self is constituted through the stories it tells. For prospective teachers, autobiographical accounts of schooling are identity-making acts, not merely recollections. Who acts in those accounts—the student-narrator or the teacher-figure—encodes a claim about who holds pedagogical power, a claim the narrator may not consciously intend.

2.2. Narrative Agency and Professional Identity

Agency in narrative has been operationalised through dependency structure (Bamberg, 1997; Wortham, 2001): the grammatical subject of transitive verbs is the agent; the object, the patient. In teacher narrative specifically, Alsup (2006) observed that prospective teachers oscillate between positioning themselves as active learners and as passive recipients of institutional authority. Beauchamp and Thomas (2009) theorised this oscillation as central to professional identity formation: teachers who never practise agential self-positioning in their stories may struggle to claim professional authority in the classroom.

At the domain level, we propose that agency allocation is not random but schema-driven: prospective teachers encode different implicit theories of 'who acts' in support situations, in didactic strategy contexts, and in assessment encounters. These domain-specific schemas are shaped by cumulative exposure during schooling and are likely to be partially shared across cohorts of similarly educated individuals—a prediction the present data allow us to test. Although our analysis focuses on agency, affect, resolution, and lexis, the same logic extends to other belief domains that teacher education seeks to influence—such as beliefs about student motivation or about classroom discipline—each of which

is likely to carry its own structural and affective signature in autobiographical memory. Because these properties index how firmly and how emotionally a belief is held, they are also what reflective practice must engage: beliefs that are affectively charged and lexically formulaic are precisely those most resistant to revision through information alone.

2.3. *Assessment as a Distinctive Narrative Domain*

The assessment literature gives particular reason to expect negative memory in this domain. Black and Wiliam's (1998) seminal review documented widespread student experiences of assessment as threatening rather than supportive. More recent work (Carless, 2007; Remesal, 2011) has shown that students' lay theories of assessment 'assessment as measurement' versus 'assessment as learning' are formed early and are highly persistent. Students who experience assessment primarily as high-stakes judgment may enter teacher education with a theory of evaluation that emphasises accuracy over formative purpose. If this theory is stored in negatively valenced narrative form, it may be particularly resistant to change through coursework alone.

Italian educational research (Castoldi, 2012; Trinchero, 2006) has documented an 'assessment culture' in Italian secondary schools characterised by relatively high stakes, infrequent formative feedback, and strong student perception of evaluative arbitrariness. The present corpus, drawn from Italian pre-service teachers, offers a test of whether this cultural context is reflected in the affective and narrative structure of autobiographical assessment memories.

3. Methods

3.1. *Research Design, Context, and Participants*

The study followed an observational, cross-sectional design with two consecutive cohorts, combining corpus-linguistic indicators with confirmatory inferential testing. No intervention was administered: the autobiographies were naturally occurring coursework artefacts rather than elicited research data. The unit of analysis was the dimension-level text (up to eight per participant) and, because these units are nested within participants, all inferential models accounted for that structure (see Section 3.4). The design is descriptive and exploratory; several of the constructs we foreground—most notably the affective and lexical distinctiveness of the Assessment domain—emerged from the data rather than from a priori hypotheses, and we therefore treat the corresponding analyses as hypothesis-generating.

With respect to research ethics, the autobiographies were produced as part of regular coursework and were pseudonymised before analysis; no personally identifying information was retained, and only aggregate, dimension-level indicators are reported. Because the study involved the secondary analysis of anonymised educational texts with no experimental manipulation, it did not require review by an institutional ethics board under the applicable national norms, and the data were processed in line with the institution's data-protection obligations.

The corpus consists of 361 pedagogical autobiographies produced by students enrolled in the General Didactic of Primary Science Formation programme at an Italian university across two academic cohorts: 2024–25 ($n = 178$) and 2025–26 ($n = 183$). This is the national teacher qualification route for Italian primary school teachers. Participants were asked to produce a structured autobiography addressing eight predetermined pedagogical dimensions (see Section 3.2). Participation was part of a compulsory course component; all data were anonymised prior to analysis. The resulting corpus comprises 2567 dimension-level text units (mean length: cohort 2024–25 = 8857 characters, $SD = 3732$; cohort 2025–26 = 10,508 characters, $SD = 7172$).

3.2. Corpus Structure

Each autobiography was structured around eight thematic prompts, corresponding to established dimensions of pedagogical competence in Italian teacher education regulations (MIUR, 2010):

(1) Support, guidance, and relational support (sostegno, guida, supporto); (2) Metacognitive teaching (didattica metacognitiva); (3) Active didactic strategies (strategie didattiche); (4) STEM/STEAM teaching (didattica STEM); (5) Humanistic teaching (didattica umanistica); (6) Classroom management (gestione della classe); (7) Learning climate (clima di apprendimento); (8) Assessment (valutazione). Participants were classified as completing all eight dimensions (2024–25: 82.6%; 2025–26: 82.0%) or as missing one or more. Analysis was conducted on all available dimension-level texts ($N_{\text{unit}} = 2567$).

3.3. Analytical Constructs

3.3.1. Student Agency Ratio (SAR)

SAR is defined as the proportion of first-person agency markers (student-as-agent) over the sum of first-person plus third-person teacher-reference markers (teacher-as-agent). First-person markers included inflected forms of the first-person singular and plural pronoun and a set of first-person autobiographical verbs (e.g., *ho imparato*, *mi sono sentita*, *ricordo*, *ho capito*). Teacher-reference markers included nominal teacher-role terms (*insegnante*, *professore/ssa*, *maestra*, *docente*) and third-person transitive constructions with teacher as subject (*ha proposto*, *spiegava*, *ci ha fatto*, *proponeva*). SAR ranges from 0 (fully teacher-agentive narrative) to 1 (fully student-agentive). SAR is deliberately a surface-level, lexicogrammatical proxy for narrative agency: it captures who is encoded as the grammatical and referential agent, rather than the fuller, multi-level positioning—of characters relative to one another, of the narrator relative to the audience, and of the narrator relative to dominant discourses—theorised by Bamberg (1997). We therefore treat Bamberg’s (1997) positioning framework as the conceptual motivation for attending to agency, while making no claim that a frequency ratio operationalises that framework in full (see Section 6.2). The binary contrast between first-person and teacher-reference markers likewise simplifies a domain in which peers, institutions, and impersonal forces may also be construed as agents.

SAR was computed at the dimension-level unit, with a minimum length threshold of 50 characters. Units below threshold were excluded. The measure is transparent and reproducible: code is available from the corresponding author.

3.3.2. Sentiment Valence

Sentiment valence was computed using a domain-specific Italian pedagogical lexicon assembled from three sources: (a) the Sentix Italian affective lexicon (Basile & Nissim, 2013); (b) polarity resources derived from SentiWordNet 3.0 (Baccianella et al., 2010) and Italian lexical mappings; and (c) domain-specific terms identified through iterative examination of the corpus, covering pedagogical efficacy (*coinvolgente*, *stimolante*, *formativo*), emotional support (*sostenuto*, *valorizzato*, *sicuro*), learning outcomes (*imparato*, *capito*, *consapevole*), and negative pedagogical experiences (*ingiusto*, *ansia*, *frustrazione*, *rigido*, *umiliato*). The final lexicon comprised 38 positive and 35 negative lemmas.

Valence was operationalised as $\text{pos_count} / (\text{pos_count} + \text{neg_count})$ per text unit, yielding a ratio from 0 (entirely negative) to 1 (entirely positive). Units with zero lexicon hits were coded as missing and excluded from valence analyses (retained for SAR and other analyses).

3.3.3. Narrative Resolution

Each dimension-level text was coded for narrative resolution through pattern-matching over a set of 14 positive and 8 negative resolution phrases (e.g., positive: *conservo*

un ricordo positivo, ricordo con entusiasmo, mi ha dato fiducia; negative: non si è risolto, ha alimentato ansia, voto ingiusto). Texts matching only positive patterns were classified ‘positive resolution’; only negative patterns, ‘negative resolution’; both, ‘mixed’; neither, ‘unclear’. The ‘unclear’ category is residual: it denotes the absence of any explicit positive or negative resolution marker and may reflect either a genuinely open-ended account or simply the absence of the specific lexical cues the procedure searches for; it should not be read as positive evidence of unresolved experience. Inter-rater reliability for this coding scheme was estimated through a 10% random sample hand-coded by two independent raters ($\kappa = 0.71$, $p < 0.001$).

3.3.4. Lexical Richness

Lexical richness was measured using Moving-Average Type-Token Ratio (MATTR; Covington & McFall, 2010) with a sliding window of 50 tokens. MATTR corrects for the well-documented negative relationship between TTR and text length, making it suitable for texts of varying length such as those in this corpus. Tokenisation used a regex-based extractor retaining Italian alphabetic forms of three or more characters, excluding punctuation and numbers.

3.3.5. School-Level Anchoring

Each dimension-level text was coded for the school level being narrated using pattern-matching over Italian school-order designations: primary (elementare, scuola primaria, maestra, and ordinal expressions such as quinta elementare); lower secondary (medie, scuola secondaria di primo grado, and ordinal expressions); upper secondary (liceo, superiori, scuola secondaria di secondo grado, istituto tecnico, and ordinal expressions). Texts matching multiple levels were assigned to the most-mentioned level; unresolvable cases were coded ‘unclear’ and excluded from school-level analyses.

3.3.6. Construct Validation Against Human Coding

To assess the validity of the two automated indicators, a stratified subsample of 160 dimension-level units (20 per pedagogical dimension), spanning the full range of automated SAR scores and both cohorts, was independently coded by two trained raters who were blind to the automated values. Each unit was rated on two five-point scales: narrative agency (1 = the teacher or another party is the agent throughout; 5 = the student-narrator is the agent throughout) and affective valence (1 = clearly negative; 5 = clearly positive). Inter-rater reliability was quantified with quadratic-weighted Cohen’s κ and the intraclass correlation [ICC(2,1)]; criterion validity was assessed as the Spearman correlation between each automated indicator and the mean of the two human ratings.

3.4. Statistical Analysis

Because each participant contributed up to eight dimension-level units, observations are nested within participants and are not independent. Overall dimension-level differences in SAR, Sentiment Valence, and lexical richness were therefore tested with the Friedman test (a repeated-measures rank test) and, to accommodate participants with incomplete dimension sets, the Skillings–Mack test; effect sizes are reported as Kendall’s W. These results were confirmed with linear mixed-effects models containing a fixed effect of dimension and a random intercept for participant, from which we also report the intraclass correlation (ICC) as an index of within-participant clustering. The focal contrast between Assessment and the other dimensions was evaluated both from the mixed-effects estimated marginal means and with paired Wilcoxon signed-rank tests (Assessment vs. each dimension, Holm-corrected), reporting rank-biserial correlations (Rosenthal, 1991) and the paired effect size Cohen’s d_z . Cross-cohort replication was assessed with a dimension $\times$ cohort

interaction term in the mixed-effects models and, per dimension, with Mann–Whitney U tests on participant-level values. Categorical outcomes (narrative resolution, school-level anchoring) were analysed with generalized estimating equations (logistic) and Cochran’s Q, alongside chi-square tests with Cramér’s V. Analyses were conducted in Python 3.12 (SciPy and statsmodels); a significance threshold of $\alpha = 0.05$ was adopted, with Holm or Bonferroni correction where multiple comparisons were made.

4. Results

4.1. Descriptive Statistics

Table 1 presents descriptive statistics for all eight dimensions. Full texts were available for 82.3% of participants across both cohorts; the remaining 17.7% had one or more missing dimension-level texts, predominantly involving the later-listed dimensions (assessment, metacognitive teaching). Missing-dimension profiles did not differ significantly between cohorts ($\chi^2 = 4.21$, $df = 7$, $p = 0.75$).

Table 1. Descriptive statistics by pedagogical dimension (N = 2567 dimension-level texts). SAR = Student Agency Ratio; Valence = sentiment valence ratio; MATTR = Moving-Average Type-Token Ratio (window = 50); % Pos./Neg. = percentage of texts with positive/negative narrative resolution.

% Neg.	% Pos.	MATTR	Valence M (SD)	SAR M (SD)	N	Dimension
2.6	28.1	0.9159	0.812 (0.317)	0.859 (0.110)	352	Support & guidance
1.5	20.3	0.9138	0.976 (0.128)	0.834 (0.160)	325	Metacognitive teaching
0.9	45.5	0.9177	0.977 (0.102)	0.789 (0.178)	323	Didactic strategies
1.6	48.9	0.9174	0.983 (0.105)	0.776 (0.195)	321	STEM teaching
1.0	47.9	0.9242	0.965 (0.139)	0.808 (0.178)	313	Humanistic teaching
1.3	18.6	0.9212	0.902 (0.214)	0.620 (0.230)	311	Classroom management
1.0	38.4	0.9245	0.937 (0.169)	0.648 (0.228)	310	Learning climate
22.8	8.3	0.8950	0.208 (0.304)	0.725 (0.231)	312	Assessment

Note. All M and SD values are means and standard deviations across dimension-level texts within each dimension. MATTR values below 0.900 indicate reduced lexical diversity.

4.2. Student Agency Ratio Across Dimensions

SAR differed across dimensions (GEE Wald $\chi^2(7) = 491.4$, $p < 0.001$; mixed-effects model LRT $\chi^2(7) = 416.4$, $p < 0.001$, ICC = 0.14; Friedman $\chi^2(7) = 365.9$, $p < 0.001$, Kendall’s W = 0.19). As Figure 1A shows, Support & guidance (M = 0.859, SD = 0.110) and Metacognitive teaching (M = 0.834, SD = 0.160) exhibited the highest SAR, indicating narratives in which the student-narrator is the more frequent grammatical agent. Classroom management (M = 0.620, SD = 0.230) and Learning climate (M = 0.648, SD = 0.228) showed the lowest SAR, indicating that the teacher-figure, rather than the student-narrator, is the more frequent agent in narratives about classroom regulation and climate.

The Assessment dimension occupied an intermediate SAR position (M = 0.725, SD = 0.231). Paired comparisons (Wilcoxon signed-rank, Holm-corrected) showed that Assessment SAR was significantly lower than the five learner-centred dimensions—Support & guidance, Metacognitive teaching, Humanistic teaching, Didactic strategies, and STEM teaching (all $p < 0.001$; rank-biserial correlations from -0.21 to -0.56)—but significantly higher than Classroom management (rank-biserial = $+0.45$, $p < 0.001$) and Learning climate (rank-biserial = $+0.30$, $p < 0.001$). Assessment therefore did not belong to the low-SAR cluster formed by the two observer-oriented dimensions; it occupied a genuinely intermediate position.

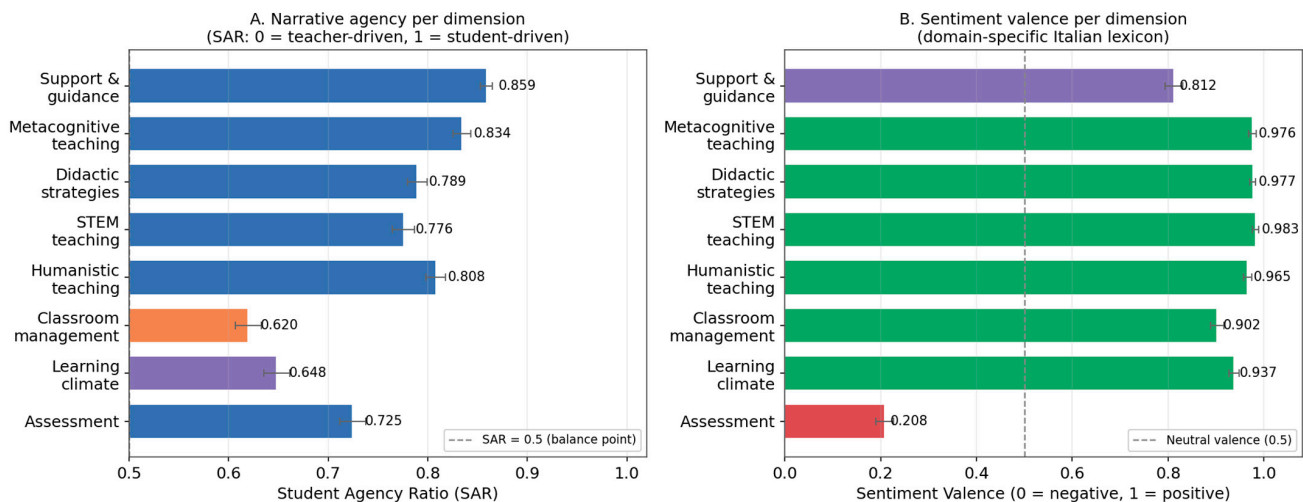


Figure 1. (A) Student Agency Ratio (SAR) and (B) Sentiment Valence by pedagogical dimension. Error bars = ± 1 SE. Dashed vertical lines indicate the SAR = 0.50 balance point and the Valence = 0.50 neutral point respectively.

4.3. Sentiment Valence Across Dimensions

Sentiment valence varied even more strongly across dimensions than SAR, and the effect was robust to the nesting of dimensions within participants (GEE Wald $\chi^2(7) = 1807.1$, $p < 0.001$; mixed-effects model LRT $\chi^2(7) = 1981.4$, $p < 0.001$, ICC = 0.005; Friedman $\chi^2(7) = 403.7$, $p < 0.001$, Kendall's $W = 0.59$). Seven of the eight dimensions had mean valence above 0.79 (range: 0.812–0.983), consistent with the predominantly positive orientation of autobiographical school memories documented in the retrospective cognition literature (Ross, 1989). Of these, Support & guidance was the lowest but still clearly positive ($M = 0.812$), reflecting some mixed accounts that narrated both supportive and emotionally difficult episodes. Only the Assessment dimension ($M = 0.208$) fell well below the positive range.

The Assessment dimension stood apart from all others as by far the most negatively valenced domain (see Table 1). Because each participant contributes one value per dimension, we quantified this gap with a within-participant effect size: the paired standardised difference between Assessment and the average of the other dimensions was very large (Cohen's $d_z = 2.31$, 95% CI [2.09, 2.54], $n = 274$), and Assessment was significantly more negative than every other individual dimension in paired comparisons (Wilcoxon signed-rank, all Holm-corrected $p < 0.001$, rank-biserial correlations from -0.95 to -1.00). The pattern did not differ between the two cohorts (Mann–Whitney U , $p = 0.12$) and is mirrored in the narrative resolution distribution (see Section 4.4): Assessment is the only dimension in which negative resolutions outnumber positive ones, a reversal not seen in any other domain. Figure 2 plots the mean SAR of each dimension against its mean Sentiment Valence, showing Assessment isolated at intermediate agency and strongly negative valence while all other dimensions cluster in the positively valenced range.

4.4. Narrative Resolution

Narrative resolution varied strongly across dimensions after accounting for clustering (GEE Wald $\chi^2(7) = 161.3$, $p < 0.001$ for negative resolution; within-subject Cochran's Q $\chi^2(7) = 304.0$, $p < 0.001$, complete cases $n = 293$). Assessment had roughly nineteen times the odds of a negative resolution compared with STEM (OR = 18.6, 95% CI [7.5, 46.0]) and eleven to thirty-one times the odds relative to the other dimensions. Figure 3 displays the full resolution distribution. STEM teaching (48.9% positive, 1.6% negative), Humanistic teaching (47.9% positive, 1.0% negative), and Didactic strategies (45.5% positive, 0.9% negative) show the highest rates of positive resolution consistent with their high sentiment valence scores.

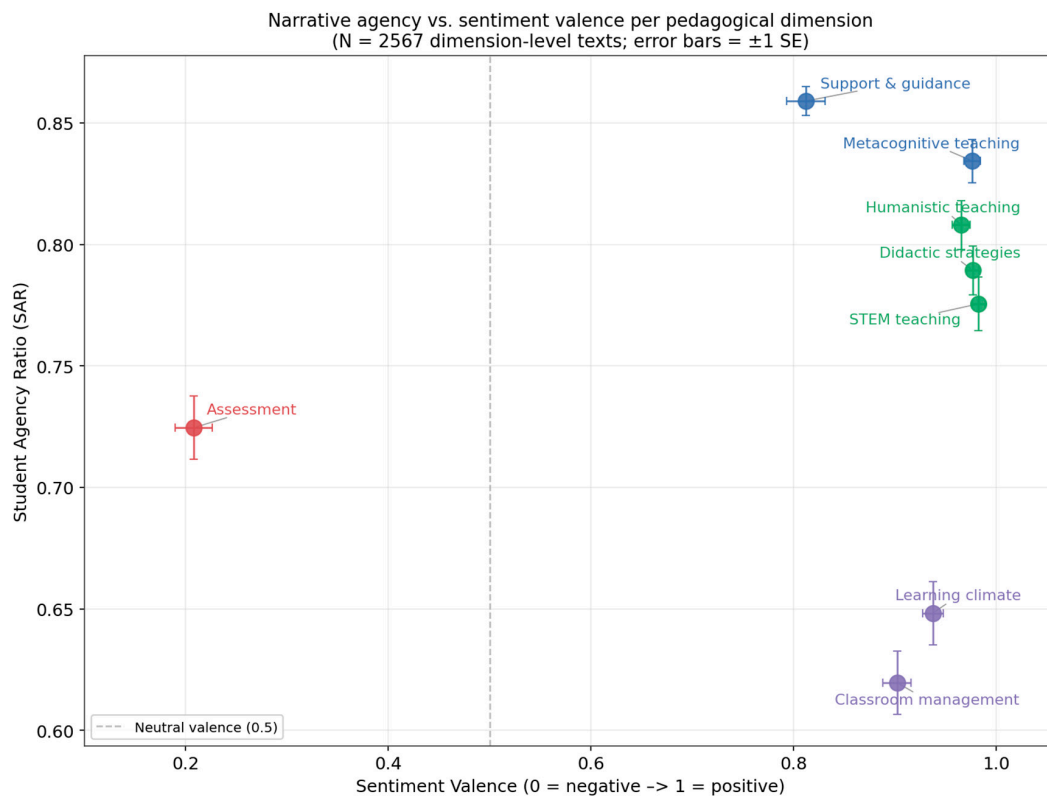


Figure 2. Bivariate distribution of Student Agency Ratio (SAR) and Sentiment Valence across pedagogical dimensions. Each point represents the dimension mean; error bars = ± 1 SE. The Assessment dimension occupies a unique region of the space: intermediate SAR, strongly negative valence.

Assessment uniquely reverses this pattern: only 8.3% of assessment texts were coded as positive resolution, compared with 22.8% negative, 2.2% mixed, and 66.7% unclear. The ‘unclear’ dominance in assessment narratives is itself informative: it reflects texts in which neither positive nor negative resolution markers are present, consistent with a narrative schema of unresolved or open-ended evaluative experience rather than a discrete episodic memory with a legible outcome. Because the ‘unclear’ code also captures texts that simply lack the specific lexical cues the procedure detects, this reading is offered tentatively.

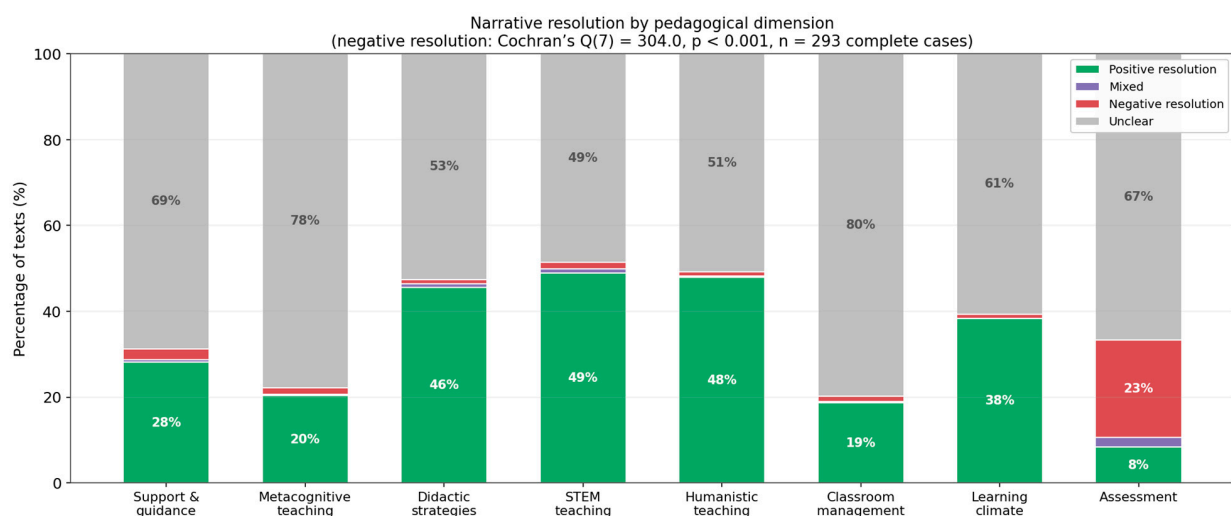


Figure 3. Narrative resolution by pedagogical dimension. Percentages are within-dimension proportions. The assessment dimension is the only one in which negative resolution substantially exceeds positive resolution (within-subject Cochran's $Q = 304.0$, $df = 7$, $p < 0.001$, complete cases $n = 293$).

4.5. Lexical Richness

MATTR scores were high and relatively uniform across seven dimensions (range: 0.9138–0.9245), indicating consistent lexical diversity in the narrative production. The Assessment dimension showed the lowest MATTR ($M = 0.8950$), significantly below all other dimensions. This finding suggests that assessment narratives draw on a more restricted and formulaic vocabulary than the other domains—rather than the more individualised narrative elaboration observed elsewhere—a convergence we interpret in the Discussion (Section 5.2).

4.6. School-Level Anchoring of Pedagogical Memories

School-level anchoring differed significantly across dimensions ($\chi^2 = 123.50$, $df = 14$, $p < 0.001$, Cramér's $V = 0.168$; Figure 4). STEM teaching showed the highest primary-school anchoring (47.6%), followed by Support & guidance (43.8%) suggesting that first encounters with science-as-experience and with relational support from teachers are predominantly associated with early schooling. This finding aligns with developmental research on science attitude formation (Archer et al., 2010) and attachment-based models of teacher–student bonding (Pianta, 1999).

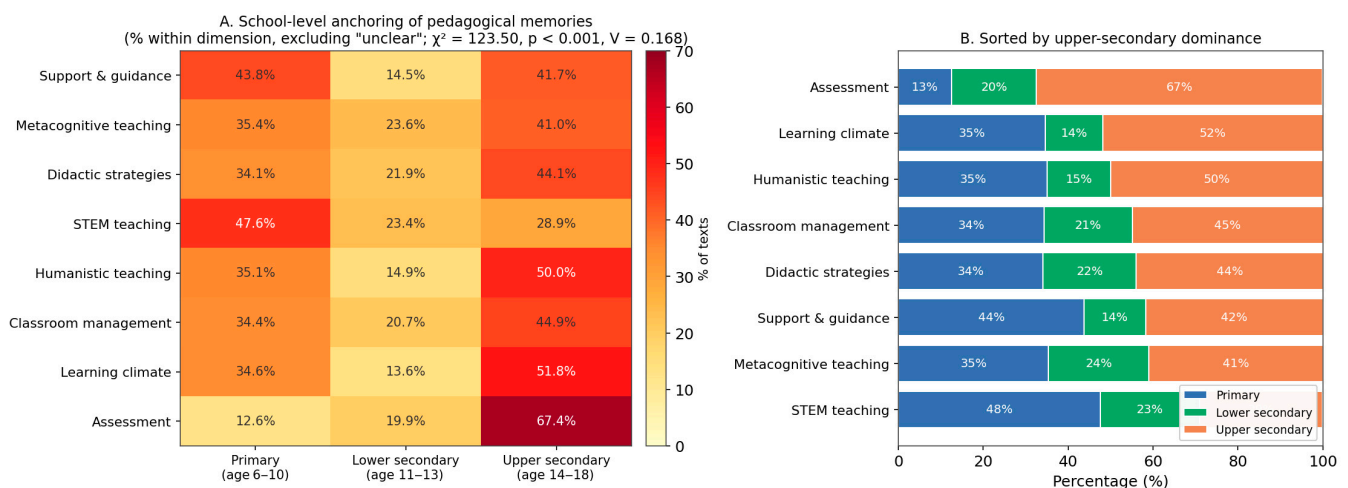


Figure 4. School-level anchoring of pedagogical memories. (A) Heatmap of within-dimension percentage distribution across primary, lower secondary, and upper secondary school levels. (B) Dimensions sorted by upper-secondary dominance. Percentages exclude texts coded as ‘unclear’ school level.

Assessment memories, by contrast, were predominantly anchored in upper secondary school (67.4%), the highest upper-secondary proportion of any dimension. Learning climate (51.8%) and Humanistic teaching (50.0%) were also upper-secondary dominant, but their valence profiles are strongly positive. The convergence of upper-secondary anchoring, negative valence, and negative resolution in the Assessment dimension points specifically to secondary-school evaluation cultures as the formative site of prospective teachers’ assessment beliefs.

4.7. Cross-Cohort Stability

To test whether the dimension-level profiles replicated across the two independently recruited cohorts, we fitted mixed-effects models with a dimension $\times$ cohort interaction (random intercept per participant). For both Sentiment Valence and SAR the dimension $\times$ cohort interaction was non-significant (valence: GEE Wald $\chi^2(7) = 9.1$, $p = 0.24$; SAR: GEE Wald $\chi^2(7) = 13.1$, $p = 0.07$; for SAR the mixed-effects LRT was marginal, $\chi^2(7) = 15.4$, $p = 0.03$), and no individual dimension differed between cohorts after correction. The agency

and valence profiles therefore replicated across cohorts; we read this as evidence of replication rather than as proof of strict equivalence, since a non-significant test cannot establish the absence of a difference. Lexical richness was the exception: its dimension $\times$ cohort interaction was significant (GEE Wald $\chi^2(7) = 17.2$, $p = 0.02$; mixed-effects LRT $\chi^2(7) = 27.4$, $p < 0.001$), with seven of the eight dimensions showing higher lexical richness in the 2025–26 cohort (all Holm-corrected $p < 0.01$) and only Assessment unchanged; the lexical-richness profile should therefore not be described as cohort-invariant. Figure 5 displays the within-dimension distributions for both cohorts.

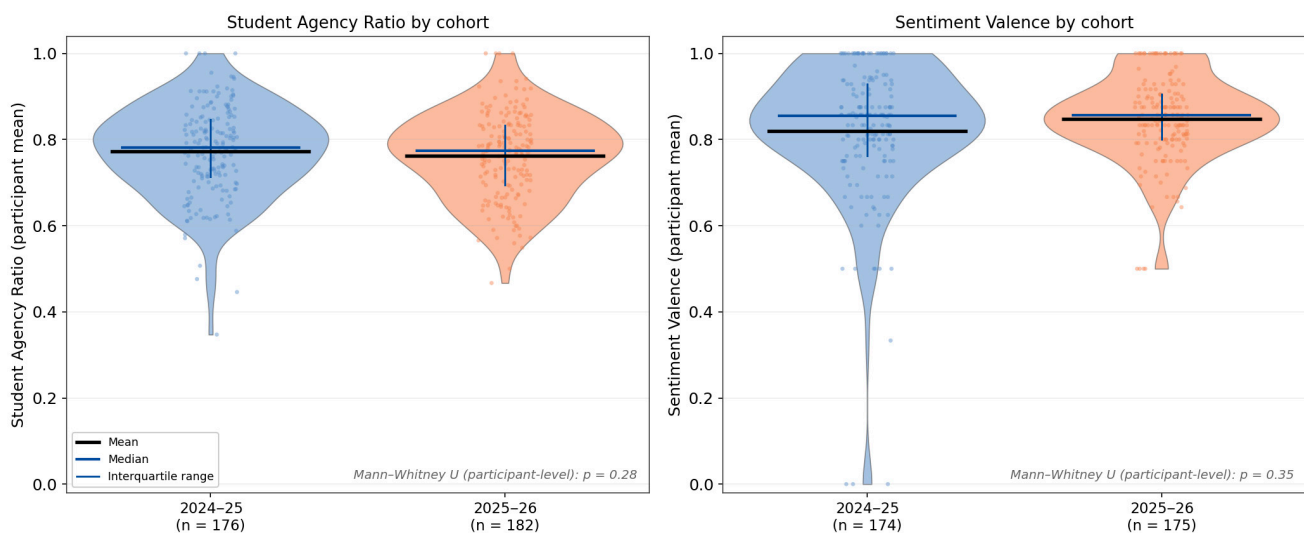


Figure 5. Cross-cohort comparison of Student Agency Ratio and Sentiment Valence, shown as participant-level means by cohort. Mann–Whitney U tests at the participant level indicate no significant cohort differences (valence $p = 0.35$; SAR $p = 0.28$).

4.8. Relationship Between SAR and Valence

Pearson correlations between SAR and Valence were computed per dimension. Correlations were small and non-significant for seven of the eight dimensions (range: $r = -0.065$ to $r = 0.040$, all $p > 0.28$), with one small but significant exception for Support & guidance ($r = +0.166$, $p = 0.005$). This near-zero relationship across dimensions indicates that agency positioning and affective valence are largely orthogonal narrative dimensions: prospective teachers can narrate high-agency experiences with negative valence (e.g., an episode in which one successfully resisted an unfair assessment), and low-agency experiences with positive valence (e.g., a teacher whose care one passively received but warmly remembers). The construct validity of SAR and Valence as distinct analytical dimensions is thereby supported.

4.9. Construct Validation

The two automated indicators showed good agreement with independent human coding (Section 3.3.6). Automated SAR correlated strongly with the mean human agency ratings (Spearman $\rho = 0.77$, 95% CI [0.70, 0.82], $n = 160$), and automated valence correlated strongly with the mean human valence ratings ($\rho = 0.71$, 95% CI [0.62, 0.78]). Inter-rater reliability was substantial for both constructs (agency: weighted $\kappa = 0.64$, ICC = 0.65; valence: weighted $\kappa = 0.74$, ICC = 0.74). Because automated valence approached its ceiling for the predominantly positive dimensions, the rank correlation understates the level of agreement; framed as a classification task, the indicators separated the units that human raters judged negative or low in agency almost perfectly (valence: area under the ROC curve = 1.00; agency: AUC = 0.94). The transparent, rule-based indicators therefore track

human judgements of agency and sentiment well on this subsample, although they do not replace the fuller relational positioning discussed in Section 6.2.

5. Discussion

5.1. *The Structured Narrative Landscape of Pedagogical Memory*

The core finding of this study is that the eight pedagogical dimensions elicited in these autobiographies do not occupy a single undifferentiated narrative space. They form a structured landscape in which agency allocation, affective valence, and narrative resolution cluster in dimension-specific configurations that are stable across two cohorts, with lexical richness a partial exception (Section 4.7). We term this configuration the pedagogical narrative schema of the dimension: a partly tacit, partly shared template that prospective teachers activate when narrating that domain of school experience. We use this term as an interpretive construct rather than a directly measured one: the schema is inferred from the converging patterns in agency, valence, resolution, lexis, and school-level anchoring, and the label should be read as a hypothesis about the underlying cognitive structure—to be tested against qualitative and validation evidence—rather than as an established empirical entity. This domain-differentiated profile contrasts with recent work on experienced, well-evaluated teachers, whose autobiographical narratives are dominated by positive affect across themes (Barbosa & Marín-Suelves, 2024); the markedly negative valence we observe for the Assessment domain among prospective teachers may therefore reflect career-stage or selection effects—positivity increasing as teachers gain experience, or as less positively oriented individuals leave the profession—an alternative that cross-sectional data cannot rule out.

This concept extends Pajares's (1992) notion of teacher beliefs by specifying the narrative structure through which beliefs are stored and retrieved. A prospective teacher who enters training does not merely 'believe' that classroom management is the teacher's responsibility: they carry a narrative in which the teacher is consistently the agent, the classroom is the setting, and the affective tone is positive, but agency is attributed to the instructor rather than to the self. This narrative has consequences for professional identity: it may reinforce a conception of classroom management as something that skilled teachers do to students, rather than something that emerges from relational co-construction.

The finding that SAR and Valence are largely orthogonal is theoretically important. It implies that the affective loading of a memory and the identity positioning it encodes operate through different mechanisms a distinction that Beauchamp and Thomas (2009) intuited theoretically but which the present data demonstrate empirically at scale. Teacher education interventions targeting belief change may therefore need to address both dimensions separately: surfacing the identity implications of who acts in one's memories is a different pedagogical task from processing the emotional residue of those memories.

5.2. *Assessment as a Structurally Distinct Narrative Domain*

The Assessment dimension demands extended discussion. Its profile—intermediate SAR, strongly negative valence, predominantly negative or unresolved narrative resolution, lowest lexical richness, and predominant anchoring in upper secondary school—is qualitatively different from all other dimensions and consistent across both cohorts.

The low lexical richness finding is perhaps the most diagnostically important. MATTR below 0.900 indicates a lexical repertoire that is narrower and more formulaic than that of the other dimensions. When prospective Italian teachers narrate experiences they construe as unfair assessment, they tend to reach for the same words: ingiusto, voto, professore, penalizzato, pregiudizio. This lexical convergence is consistent with assessment-related school memories being stored as instances of what may function as a shared cultural

script—what we might call a *perceived assessment-injustice narrative* rather than as fully individualised autobiographical episodes. Cultural scripts of this kind are precisely what Schank and Abelson (1977) predicted would be most resistant to modification through experience, and what Pajares (1992) identified as requiring explicit, confrontational pedagogical intervention to revise—an intervention that, in the terms of reflective practice, must reach the critical level at which the underlying frame of reference is questioned rather than merely described (Hatton & Smith, 1995; Mezirow, 1991).

The school-level anchoring of assessment memories in upper secondary school is also theoretically generative. It maps onto the Italian education system's documented intensification of high-stakes evaluation at the liceo and istituto tecnico levels (Castoldi, 2012), and particularly onto the maturità examination experience that virtually all participants will have undergone within five years of writing these accounts. The emotional intensity and evaluative salience of that experience likely contribute to its dominant encoding in assessment memory. Curriculum designers might consider that targeting assessment beliefs in pre-service training effectively means engaging with memories that are recent, emotionally charged, and culturally shared—a combination that requires deliberate autobiographical disruption rather than information provision.

6. Conclusions

Prospective teachers arrive at initial training carrying structured, domain-differentiated narrative schemas for each major area of pedagogical competence. These schemas—identified here through Student Agency Ratio, Sentiment Valence, narrative resolution, and lexical richness—are largely stable across cohorts (with lexical richness a partial exception) and reproducible through computational text analysis of Italian autobiographical accounts. The Assessment domain is structurally unique: it concentrates negative affect, formulaic language, what may function as a shared cultural script of perceived evaluative injustice, and memories anchored in upper secondary school. By describing the narrative material on which reflection must operate, this approach offers reflective teacher education a more precise starting point: taking these schemas seriously, rather than assuming they can be dissolved through information alone, better positions prospective teachers to undertake the difficult, critically reflective task of revising deeply held, affectively charged, and culturally reinforced beliefs about what teaching and evaluation are for. Although the present analysis focused on written autobiographies, the same computational lens could be applied to other reflective tools used in teacher education—reflective journals, video-stimulated recall commentaries, and structured practicum reports—to test whether the domain-differentiated schemas observed here generalise across genres. Longitudinal and in-service designs that follow teachers from initial training into practice would further clarify whether these narrative schemas persist, evolve, or are reshaped by classroom experience. Because these patterns were documented in a single Italian institution and without demographic differentiation, they should be read as specific to this context rather than as established features of prospective teachers in general; multi-institution and demographically differentiated replication is needed before any broader generalisation.

6.1. Implications for Teacher Education

Three practical implications follow from these findings; because the study is observational and cross-sectional, we frame them as possible directions for teacher education to explore rather than as effects the data establish. First, the dimension-specific SAR profile suggests that prospective teachers' professional identity work should vary by domain. In classroom management and learning climate—the lowest-SAR dimensions—teacher education might explicitly invite prospective teachers to renarrate their memories from the

teacher's perspective, practising the agentic positioning they will need to enact professionally. Such guided renarration is itself a reflective-practice technique, in that it makes the original framing available for critical re-examination (Larrivee, 2000). In support and metacognitive teaching—the highest-SAR dimensions—the risk is different: prospective teachers already position themselves as the agents of their own learning, which may produce an overconfident illusion of competence (Kruger & Dunning, 1999) that structured reflection on the teacher's contribution can correct.

Second, the assessment narrative profile calls for specific, targeted intervention. The prevalence of the perceived assessment-injustice narrative, its formulaic lexical character, and its upper-secondary anchoring suggest that general 'assessment literacy' instruction will be insufficient. Effective intervention likely requires autobiographical confrontation: creating structured opportunities for prospective teachers to re-examine their specific perceived assessment-injustice memories, identify the implicit evaluation theories they encode, and encounter counter-narratives of formative assessment practice. The well-established use of 'critical incidents' as a reflective-practice method in teacher education (Tripp, 1993; Richards & Farrell, 2005) provides methodological resources for this kind of work.

Third, the cross-cohort stability of the agency and valence profiles suggests that these narrative schemas are not shaped by local or year-specific factors but by the structure of Italian secondary schooling itself. This places the locus of intervention upstream of initial teacher education—in the assessment cultures of schools—while simultaneously affirming that initial training must take the schemas as its starting point rather than assuming they can be dissolved through information alone.

6.2. Methodological Contributions and Limitations

This study demonstrates that rule-based computational analysis of Italian pedagogical autobiography, grounded in established linguistic resources, can produce theoretically interpretable results at corpus scale. The choice of rule-based over transformer-based methods was deliberate: the former is fully transparent, replicable without specialised hardware, and not dependent on external model availability—properties that support the methodological verifiability increasingly required by leading educational journals.

Several limitations should be noted. The SAR and Valence measures are necessarily approximations. SAR in particular indexes only lexico-grammatical agency and does not capture the higher-order, relational positioning central to Bamberg's (1997) framework; it should be read as a transparent, scalable proxy rather than a complete operationalisation of narrative agency. The Italian lexicon, while domain-adapted, was validated against independent human coding on a stratified subsample rather than across the entire corpus (Section 3.3.6), and the sentiment procedure does not model negation, irony, or intensity; texts with no lexicon matches were excluded, which may introduce selection bias. SAR showed strong criterion validity against hand-coded agency on the same subsample (Section 3.3.6), but it still does not capture passive or impersonal constructions. Participation formed part of a compulsory course, raising the possibility of social-desirability effects, and no participant demographics (e.g., age, subject specialism, gender) were collected, which precludes subgroup analysis. The school-level coding is pattern-based and potentially confounded in texts describing multiple school levels. The corpus is drawn from a single Italian university and may not be representative of the national prospective primary teachers' population. A further consideration is statistical non-independence: the dimension-level units are nested within participants (up to eight per person), so the original Kruskal–Wallis and Mann–Whitney omnibus tests, which assume independent observations, are not appropriate for these comparisons. We therefore re-estimated all dimension

and cohort effects with methods that respect this structure—Friedman and Skillings–Mack tests and linear mixed-effects models with a random intercept per participant—and the substantive conclusions were unchanged. The intraclass correlation attributable to participants was negligible for valence (ICC = 0.005) and modest for SAR (ICC = 0.14) and lexical richness (ICC = 0.25), indicating that the non-independence had little practical effect, especially for valence. The one construct that was not fully stable across cohorts was lexical richness, which showed a significant dimension $\times$ cohort interaction; we report this rather than treating the measure as cohort-invariant. We nonetheless recommend that future work continue to report multilevel or participant-level robustness analyses. Future work might apply transformer-based models (e.g., UmBERTo or ALBERTo) when network conditions permit, and extend the present construct validation with structured qualitative coding of larger subsamples.

Author Contributions: Conceptualization, A.R.; Methodology, G.B.; Software, G.B.; Formal analysis, G.B.; Resources, A.R.; Data curation, G.B.; Writing—original draft, A.R.; Writing—review & editing, G.B.; Supervision, A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available from the corresponding author on reasonable request. The data are not publicly available because they consist of personal autobiographical writing produced by identifiable participants.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Achinstein, B., & Ogawa, R. T. (2006). (In)fidelity: What the resistance of new teachers reveals about professional principles and prescriptive educational policies. *Harvard Educational Review*, 76(1), 30–63. [\[CrossRef\]](#)
- Alsup, J. (2006). *Teacher identity discourses: Negotiating personal and professional spaces*. Routledge.
- Archer, L., Dewitt, J., Osborne, J., Dillon, J., Willis, B., & Wong, B. (2010). “Doing” science versus “being” a scientist: Examining 10/11-year-old schoolchildren’s constructions of science through the lens of identity. *Science Education*, 94(4), 617–639.
- Baccianella, S., Esuli, A., & Sebastiani, F. (2010). SentiWordNet 3.0: An enhanced lexical resource for sentiment analysis and opinion mining. *Proceedings of LREC*, 7, 2200–2204.
- Bamberg, M. (1997). Positioning between structure and performance. *Journal of Narrative and Life History*, 7(1–4), 335–342. [\[CrossRef\]](#)
- Barbosa, M. L. d. O., & Marín-Suelves, D. (2024). Content and sentiment analysis of autobiographical narratives of experienced and well-evaluated teachers in Spain. *Education Sciences*, 14(6), 642. [\[CrossRef\]](#)
- Basile, V., & Nissim, M. (2013, June 14). *Sentiment analysis on Italian tweets*. 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (pp. 100–107), Atlanta, GA, USA.
- Beauchamp, C., & Thomas, L. (2009). Understanding teacher identity: An overview of issues in the literature and implications for teacher education. *Cambridge Journal of Education*, 39(2), 175–189.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74. [\[CrossRef\]](#)
- Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International*, 44(1), 57–66.
- Castoldi, M. (2012). *Valutare a scuola: Dagli apprendimenti alla valutazione di sistema*. Carocci.
- Clandinin, D. J. (2013). *Engaging in narrative inquiry*. Left Coast Press.
- Connelly, F. M., & Clandinin, D. J. (1990). Stories of experience and narrative inquiry. *Educational Researcher*, 19(5), 2–14. [\[CrossRef\]](#)
- Covington, M. A., & McFall, J. D. (2010). Cutting the Gordian knot: The moving-average type-token ratio (MATTR). *Journal of Quantitative Linguistics*, 17(2), 94–100. [\[CrossRef\]](#)
- Crowe, F., & McGarr, O. (2022). An investigation of preservice teachers’ apprenticeship of observation through a lens of autobiographical memory. *Journal of Teacher Education*, 73(4), 410–423. [\[CrossRef\]](#)
- Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. D. C. Heath.

- Flores, M. A., & Day, C. (2006). Contexts which shape and reshape new teachers' identities: A multi-perspective study. *Teaching and Teacher Education*, 22(2), 219–232. [[CrossRef](#)]
- Hatton, N., & Smith, D. (1995). Reflection in teacher education: Towards definition and implementation. *Teaching and Teacher Education*, 11(1), 33–49. [[CrossRef](#)]
- Korthagen, F., & Vasalos, A. (2005). Levels in reflection: Core reflection as a means to enhance professional growth. *Teachers and Teaching: Theory and Practice*, 11(1), 47–71. [[CrossRef](#)]
- Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121–1134. [[CrossRef](#)] [[PubMed](#)]
- Larrivee, B. (2000). Transforming teaching practice: Becoming the critically reflective teacher. *Reflective Practice*, 1(3), 293–307. [[CrossRef](#)]
- Lortie, D. C. (1975). *Schoolteacher: A sociological study*. University of Chicago Press.
- McAdams, D. P. (1993). *The stories we live by: Personal myths and the making of the self*. William Morrow.
- Mezirow, J. (1991). *Transformative dimensions of adult learning*. Jossey-Bass.
- MIUR. (2010). Decreto 10 settembre 2010, n. 249. Regolamento concernente la definizione della disciplina dei requisiti e delle modalità della formazione iniziale degli insegnanti della scuola dell'infanzia, della scuola primaria e della scuola secondaria di primo e secondo grado. Ministero dell'Istruzione, dell'Università e della Ricerca.
- Mortari, L. (2009). *Aver cura di sé*. Bruno Mondadori.
- Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research*, 62(3), 307–332. [[CrossRef](#)]
- Pianta, R. C. (1999). *Enhancing relationships between children and teachers*. American Psychological Association.
- Remesal, A. (2011). Primary and secondary teachers' conceptions of assessment: A qualitative study. *Teaching and Teacher Education*, 27(2), 472–482. [[CrossRef](#)]
- Richards, J. C., & Farrell, T. S. C. (2005). *Professional development for language teachers*. Cambridge University Press.
- Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In J. Sikula (Ed.), *Handbook of research on teacher education* (2nd ed., pp. 102–119). Macmillan.
- Ricoeur, P. (1991). Narrative identity. *Philosophy Today*, 35(1), 73–81. [[CrossRef](#)]
- Rosenthal, R. (1991). *Meta-analytic procedures for social research* (Rev. ed.). Sage.
- Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, 96(2), 341–357. [[CrossRef](#)]
- Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals and understanding*. Erlbaum.
- Schön, D. A. (1983). *The reflective practitioner: How professionals think in action*. Basic Books.
- Trinchero, R. (2006). *Valutare l'apprendimento nell'e-learning*. Erickson.
- Tripp, D. (1993). *Critical incidents in teaching: Developing professional judgement*. Routledge.
- Wortham, S. (2001). *Narratives in action*. Teachers College Press.
- Zannini, L. (2005). *La tutorship nella formazione degli adulti*. Guerini.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.